Handling EEG Artifacts and Searching Individually Optimal Experimental Parameter in Real Time: A System Development and Demonstration

Guang Ouyang[a], Joseph Dien[b], Romy Lorenz[c,d,e]

[a]Faculty of Education, The University of Hong Kong, Hong Kong

[b]Department of Human Development and Quantitative Methodology, University of Maryland, College Park

[c]MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK

[d]Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[e]Department of Psychology, Stanford University, Stanford, US

Contact information for corresponding author:

Address correspondence to: Dr. Guang Ouyang, Faculty of Education, The University of Hong Kong, Pokfulam, Hong Kong Island, Hong Kong, e-mail: ouyangg@hku.hk

Abstract: *Objective.* Neuroadaptive paradigms that systematically assess ERP features across many different experimental parameters have the potential to improve the generalizability of ERP findings and may help to accelerate ERP-based biomarker discovery by identifying the exact experimental conditions for which ERPs differ most for a certain clinical population. Obtaining robust and reliable ERPs online is a prerequisite for ERP-based neuroadaptive research. One of the key steps involved is to correctly isolate EEG artifacts in real time because

they contribute a large amount of variance that, if not removed, will greatly distort the ERP obtained. Another key factor of concern is the computational cost of the online artifact handling method. This work aims to develop and validate a cost-efficient system to support ERP-based neuroadaptive research. *Approach.* We developed a simple online artifact handling method, single trial PCA-based artifact removal (SPA), based on variance distribution dichotomies to distinguish between artifacts and neural activity. We then applied this method in an ERP-based neuroadaptive paradigm in which Bayesian optimization was used to search individually optimal inter-stimulus-interval (ISI) that generates ERP with the highest signal-to-noise ratio. *Main results.* SPA was compared to other offline and online algorithms. The results showed that SPA exhibited good performance in both computational efficiency and preservation of ERP pattern. Based on SPA, the Bayesian optimization procedure was able to quickly find individually optimal ISI. *Significance.* The current work presents a simple yet highly cost-efficient method that has been validated in its ability to extract ERP, preserve ERP effects, and better support ERP-based neuroadaptive paradigm.

Introduction

Event-related potentials (ERPs) are one of the most studied phenomena in cognitive neuroscience research with electroencephalography (EEG). Many different ERPs components have been studied (e.g., P300, MMN, ERN etc.) and linked to various cognitive processes and states. Moreover, ERPs have gained traction in the clinical field as they differ in patients suffering from neurological and psychiatric conditions as well as from disorders of consciousness (e.g., Morlet et al., 2014; Näätänen et al., 2012); and thus, they may yield clinically relevant biomarkers in the future.

However, at the same time, it has been pointed out that the scalability of research into mind-brain relationships is hindered by the gap between the enormous complexity of neurocognitive systems and the simplicity of laboratory-based controlled experimentation (Chen & Pesaran, 2021; Nastase, et al., 2020). This is also the case for ERP-based research that is often conducted in strictly controlled lab environments by using often lab- or study-specific experimental paradigms and settings. Many decisions about the experimental paradigm (e.g., type of stimuli, number of stimuli, number of distractors, stimulus duration, inter-stimulus interval etc.) have been made relatively ad hoc by the experimenters (following some heuristics) without systematically assessing their effect on ERP characteristics. Consequently, it is not well understood how replicable ERP findings are in general, how generalizable ERP-based findings are to a much broader class of experimental stimuli or experimental setting (for a more general discussion of this problem in cognitive neuroscience, see Yarkoni et al. 2019), or under which experimental condition certain ERP characteristics emerge or break down. That this is a timely issue is reflected in the recent "#EEGManyLabs" initiative (Pavlov et al. 2021) where many different labs are trying to replicate influential ERP findings by using the originally reported experimental paradigms across all replication attempts. While this is a very important step forward for the field, this initiative cannot address questions about the generalizability of ERP findings more broadly.

A conceptual solution is to develop closed-loop experimental systems that are capable of updating experimental parameters, formulate and test hypotheses in real-time (Chen & Pesaran, 2021; Krol et al., 2020; Lorenz, et al., 2017). From a broader perspective, closed-loop experimental systems have a long tradition not just in different fields of neuroscience (Chen & Pesaran, 2021; DiMattina et al., 2013), but also in other disciplines in general (Watson et al., 1983; Settles, 2009; Sutton et al., 1998). Equally, neuroadaptive research paradigms involving

real-time EEG have a long history (Vidal, 1977), particularly in the field of brain-computer interfacing (BCI; Lotte et al., 2018a, 2018b; Roc et al., 2020; Zander et al., 2016). However, in neuroadaptive BCI research, the primary goal has been on optimizing the control ability of users (Lotte et al., 2018a, 2018b) or enriching human-machine interaction more broadly (Zander et al., 2016; Krol et al., 2018). Applying neuroadaptive paradigms in the context of automatically searching for optimal experimental parameters for addressing basic neurocognitive research questions involving human participants (e.g., which inter-stimulus-interval is optimal for hosting a specific neural effect) has been rare to date (but see Costa et al., 2021).

To be more specific, by leveraging online EEG data analysis in combination with intelligent optimization algorithms (e.g., Bayesian Optimization), many more experimental parameters can be explored in a single experimental session with the aim to systemically investigate which experimental parameters drive specific ERP effects. In principle, this methodology would boost the efficiency of basic research in neural and cognitive mechanisms and reduce experimenter bias at the same time (Lorenz, et al., 2017). The implementation of this concept has been carried out recently, with predominant focus on studies involving functional magnetic resonance imaging (Lorenz et al., 2016, 2018). Recent studies also highlighted the potential of the approach for accelerating biomarker discovery in stroke patients and for tailoring non-invasive brain stimulation parameters to the individual (Lorenz et al., 2019, 2021).

Here, we argue that it would be beneficial to develop neuroadaptive research paradigms that integrate a functional module of online experimental parameter optimization for basic neurocognitive research. However, many fundamental questions about this topic remain unclear. For example, are the usually subtle variations in ERPs in basic research paradigms obtained online able to drive an online optimization module in a practical sense? The worst scenario is

that the ERPs obtained from a relatively short duration are not a clearly structured function (e.g., too noisy) of a factor of interest from which an optimum can be reliably detected, thus it may not be practical to build an online ERP-based optimization loop. The present work was dedicated to investigating this issue.

A crucial step in developing an ERP-based neuroadaptive research paradigm is to evaluate the feasibility of obtaining high-quality ERPs with sufficient signal-to-noise ratio in real-time, which entails an advanced procedure of handling online noise and artifacts. As is well known, electroencephalographic (EEG) data contain a large amount of artifacts that can overwhelm genuine neural activity. For instance, eye blink artifacts can generate an amplitude easily surpassing 200 μV, while a typical ERP component is less than 20 μV (Plochl, et al., 2012; Berg, et al., 1994). Neural effects, i.e., the difference between conditions or groups, can be even smaller—commonly within a few microvolts. Such an effect size would normally require a couple dozen participants for adequate statistical power even without the presence of major artifacts; therefore, separating out large artifacts is crucially important. To obtain the purest possible neurophysiological data from the EEG, researchers have identified various non-neural sources that contribute to the variance of EEG data. They include ocular artifacts, muscle activity, heartbeats, line noise, electrode connectivity-related noise, and so on. Although a perfect correction of all artifactual components is theoretically impossible, and oftentimes the application is context-specific (Jiang, et al., 2019; Mannan et al., 2018; Urigüen et al., 2015), very advanced algorithms including those incorporating deep-learning neural networks (Pion-Tonachini, et al., 2019) have been developed to effectively identify and isolate them with high generalizability. However, in an neuroadaptive paradigm that requires online extraction of relevant neural information, computational cost (or computational simplicity) of artifact handling

is an important factor of concern. Many studies opted to apply a simple amplitude threshold-based method to discard epochs with abnormally high amplitudes (e.g., da Costa et al., 2021; Kangassalo et al., 2020). This approach is computationally low-cost, but it may sacrifice too much information, and those retained data segments may still contain artifactual components.

Artifacts caused by different sources may have very different magnitudes. For example, heartbeat artifacts generate much less variance than ocular artifacts (Park, et al., 2019). Most of the current artifact-handling methods aim at capturing all kinds of non-neural components, regardless of the variance they contribute to the EEG data. A general guidance would recommend removing all of the detected artifacts as long as they are unambiguously identified as an artifact. To unambiguously isolate and identify various artifacts, a high degree of sophistication is involved in most current methodologies. First, high-level statistical features and relationships need to be modeled and identified to differentiate and decompose different sources (e.g., Delorme, et al., 2007). Second, sophisticated multi-dimensional feature extraction and classification algorithms would then be applied (Pion-Tonachini, et al., 2019; Winkler, et al., 2011, 2014). The algorithms involved in these two steps are usually computationally demanding, so they are usually applied offline.

Neuroadaptive research paradigms require online handling of artifacts. In these paradigms, sophisticated but time-intensive offline methods, especially those relying on extracting long-period statistical features, are not applicable. Conversely, online procedures are optimized for speed over accuracy to a greater degree. However, the majority of the current methodologies for online artifact handling in the BCI field is still based on feature extraction and classification algorithms that are computationally costly. A concise overview is given here: (1) The FORCe method (Daly, et al., 2015) decomposes a short EEG segment into the wavelet space and applies

ICA to decompose the wavelet coefficients. Wavelet transformation is a computationally heavy procedure (heavier than Fourier transform) in real time processing. The independent components analysis (ICA) algorithm used in FORCe, second order blind identification (SOBI) (Belouchrani, et al., 1997) is also computationally heavier than PCA because SOBI conducts matrix diagonalization over multiple time lags. The classification module in FORCe is also computationally heavy as it calculates many different aspects of statistical feature for identification of artifact components. (2) Methods that are based on empirical mode decomposition (EMD) (Andrade, et al., 2006) rely on an iterative identification of extrema in the data, which is computationally heavier than analytic methods that can be implemented by matrix operation. (3) ICA can also be applied in an online mode by applying a de-mixing matrix derived from an existing dataset (training data) to online data segments. The ICs identified as artifacts from the training data are removed and the non-artifact ICs are back-projected to the scalp EEG. However, the online data may contain novel artifact features (e.g., channel noise) that were not present in the training data and thus would not be captured in online processing. (4) Some online methods were particularly designed for removing ocular artifacts based on a prior knowledge of ocular artifact features (Egambaram, et al., 2019; Kobler, et al., 2020; Nguyen et al., 2012; Somers et al., 2016); however, only removing ocular artifacts would not be sufficient for an ERP-based neurocognitive study because there are many other high-amplitude artifacts (e.g., due to drastic body movements). In sum, most of the existing methods rely on two computational modules that are relatively costly. One is the characterization of high-order or complex features of artifacts and the other one is the online classification techniques applied on those features.

An exception is the method of artifact subspace reconstruction (ASR) (Mullen et al., 2015; de Freitas, et al., 2020; Kothe et al., 2016). ASR is an online artifact handling method that is based

on a relatively simple but practical algorithm. ASR identifies artifacts in the subspace dimensions separately, which are defined by the loadings of a single principal component (PC) derived from a segment of online data. In each subspace, if the component amplitude surpasses a threshold, the component will be dropped, and the cleaned data are reconstructed from the remaining components. The thresholds for different PC subspaces are different and are determined by calibration data. Computationally, ASR is very low-cost as it is based purely on simple matrix operations and numerical thresholding. The computational cost is substantially lower than that of online extraction and classification of complex features. ASR has been shown to be able to effectively isolate artifacts in real time (Mullen et al., 2015).

In this article, we present a new and simpler approach that is more cost-efficient and may be more suitable for the specific application in ERP-based neuroadaptive research. We propose that there is room to further and substantially reduce the computational cost and simplify the algorithm when the application domain is ERP-based research. Similar to ASR, the core computation is based on PCA, but the algorithms for identifying artifact components and online data processing in the following aspects may be changed to further reduce computational cost. First, the module of calibration data may be removed. Second, in an ERP-based paradigm, only segments encompassing events need to be processed, thus the computational cost can be further reduced by processing less data and discarding the step of smoothing adjacent segments. Third, the procedure of thresholding may be further simplified by a simple cut-off on the PCs obtained online, rather than calculating a real-time threshold incorporating the PC configurations from the calibration data.

We further propose a refinement that is specific to ERP-based neuroadaptive studies. In such a research paradigm, the ERPs are averaged from multiple trials in a recent segment of online

EEG. For example, one could obtain the ERPs for rare and frequent trials in a recent segment of online EEG in an oddball paradigm and use the ERP difference as a neural indicator of novelty response to assess mental states. Because an average step is involved, it can be expected that small-variance artifacts are effectively cancelled out if they are not time-locked to the event of interest. Besides, the effectively averaged-out small-variance artifact affects ERP effects (e.g., amplitude difference between conditions) only to a minor degree. Therefore, an algorithm can be oriented to capture large-variance artifacts, which would bear a substantially lower computational complexity than those that are based on high-level statistical features (see the review above). The computational cost saved by simple algorithms can then be used for other modules such as extraction of complex neural activity (ERP) features, data visualization, optimization algorithms, and so on.

In sum, we propose that applying methods that are oriented toward large-variance artifacts may be sufficient, and may greatly improve computational cost-efficiency for ERP-based neuroadaptive research; however, a major remaining question is "how large is large?" It would require some qualitatively distinguishable features in the variance distribution pattern of EEG constituent components for developing a theoretically sound method of isolating artifacts based on variance contributions. We propose that it is highly likely to observe a bimodal pattern in the variance distribution of EEG constituent components based on the observation that most large-amplitude artifacts are substantially larger than the level of neural activity. Such a bimodal pattern could serve as a basis for developing a simple and practical procedure of fast artifact removal.

This article presents a simple method, single trial PCA-based artifact removal or SPA, that utilizes the difference in amplitude between major artifacts and neurophysiological signals to

clean EEG data in an online mode. SPA is not intended to remove all artifacts, but rather to efficiently remove major large-variance artifacts in order to obtain a brain response pattern with sufficient accuracy for ERP-based neuroadaptive research paradigms; therefore, it is crucially important to know if this online method is able to preserve the pattern of ERP and subtle neural effects as compared to the major offline method, ICA (Jung et al., 2000), while having a superior performance in computational efficiency. We evaluate the performance of this method in both computational efficiency and preservation of ERP patterns and neural effects in a newly collected EEG dataset with 200 participants in a visual oddball task and an auditory oddball task by comparing it with other offline and online methods. After the comparative evaluation, we set up an online ERP-based neuroadaptive paradigm that aimed to find an individually optimal inter-stimulus-interval parameter that generates the largest signal-to-noise ratio of ERP for a face judgment task. The online optimization was based on Bayesian optimization. SPA was applied to this neuroadaptive paradigm to remove artifact online before the optimization module.

Materials and Methods

Participants. A dataset of 200 participants (62 males, 138 females, 18-40 years old, mean: 25.1, SD: 4.5) performing visual and auditory oddball tasks was used to evaluate the performance of the artifact handling methods. Five additional participants (4 males, 1 female, 21-34 years old, mean: 28.8, SD: 5.1) were recruited for the neuroadaptive face judgement task. The participants were all healthy adults with normal or corrected-to-normal vision and without any neurological disorder. The study was approved by the ethics committee in The University of Hong Kong (Ref no.: EA1901017). The participants signed a written consent form before the data recording. Each participant received 100 Hong Kong dollars per hour as a compensation for participating the experiment.

Tasks for evaluating the artifact handling methods. The participants were instructed to perform a visual oddball task and an auditory oddball task. During the task, the participant sat on a chair with a visual distance of about 60 cm to the monitor. Chin rest was not used. The participants were instructed to minimize their physical movement. In the visual oddball task, the participants watched a sequence of 160 color squares (blue: 135, red: 24, yellow: 1) presented one-by-one on the screen with a duration of 200 ms. The task was to count how many different colors were in the sequence (the participants did not know how many in advance). The blue and red were counterbalanced across participants, i.e., for half of the participants, it was: blue: 24, red: 135, yellow: 1. Only blue and red squares were used in data analyses, serving as frequent and rare conditions, respectively. The inter-stimulus-interval (ISI) was uniformly distributed between 1700 ms and 2700ms. In the auditory task, the parameters were the same as for the visual oddball task except that the three colors were replaced by three tones (400 Hz, 600 Hz, 650 Hz) and the ISI was uniformly distributed between 1200 ms and 1700 ms.

EEG collection and basic pre-processing. EEG data were collected in a sound-attenuated room using Brain Product's actiCHamp amplifier with 32 channels referenced to a reference internal to the amplifier. The following pre-processing steps were conducted on the data: 1) down sampling to 150 Hz; 2) bandpass filtering between 1 and 40 Hz; 3) re-referencing to the average reference (Dien, 1998; Bertrand, et al., 1985). After these basic preprocessing steps, the data still contained a large amount of artifacts. The three following methods were then applied to this preprocessed dataset:

1) Independent Component Analysis. ICA is a well-established method for handling EEG artifacts (Jung, et al., 2000). Here, ICA was applied on the whole dataset covering the entire task because the algorithm requires large number of sampling points to model statistical

independence (thus, it was mostly used as an offline method). The results of the ICA method served as a benchmark for the two fast methods. The extended Infomax ICA algorithm in the EEGLAB toolbox (Delorme, et al., 2004; Lee et al., 1999) was used in the present work (The scripts for this study are available online at https://github.com/guangouyang/spa). This ICA algorithm treats channels as variables and time points as observations. After ICA, the Multiple Artifact Rejection Algorithm (MARA) toolbox (Winkler, et al., 2011) for automatically identifying and removing artifacts was applied (probability threshold was set to be 0.5).

2) Single Trial PCA-based Artifact removal. The SPA algorithm consists of the following procedure. First, varimax spatial PCA was applied on each EEG segments surrounding each time marker of stimulus onset. The time window was set as from -200 ms to +800 ms with respect to the stimulus onset. After applying PCA, the amplitude for each PC was calculated as square root of the PC variance. The PCs with an amplitude larger than T were truncated and the rest of the PCs were back-projected to the sensor space, such that a cleansed segment was obtained. This cleansed segment is a single trial that is used to generate the average ERP. The parameter T is adjustable and was set as 30 in the present application based on the observation of the data that the value 30 well covers the base cluster in the PC amplitude distribution. Given the stability in the magnitude range of major artifacts (e.g., ocular), this threshold parameter is expected to be a fixed one that can be universally applied across experiments within the same lab setting. In the neuroadaptive experiment later, we applied the same threshold value.

Clustering of the artifact components removed by SPA based on the scalp map pattern. To examine the scalp map pattern of the PC components removed by SPA and validate their artifact nature, we conducted a clustering of all the scalp maps of SPA-removed PCs according to the following procedure. 1) The scalp maps of all PCs identified as artifact from all single trials and

participants were pooled together. 2) The pair-wise correlations amongst all the scalp maps were calculated. 3) The scalp map that has the largest mean correlation with all others were used as the first seed, and all the other maps that have a correlation value higher than 0.2 with the seed map were selected as the first cluster and the average map of them was obtained. 4) The first cluster was removed from the correlation matrix and 2) and 3) were repeated to get the second cluster and the procedure goes on and on. The clustering was applied to both of the two tasks and the average patterns of the first ten clusters as well as their percentages were shown.

3) Artifact Subspace Reconstruction. ASR was developed for online artifact removal by Mullen and colleagues (Mullen, et al., 2015). ASR also uses a PCA core to handle artifacts. First, a reference data segment that contained very few artifacts (e.g., from a resting state recording without much movement) was collected. Second, PCA was applied to the covariance matrix obtained as the median of the covariance matrices from separate segments in the reference data. The median was used because it is robust to outliers; that is, even if the reference data contains a few large-variance artifacts, it would not greatly affect the resultant median covariance matrix. The calibration data were then projected onto the PC space, and ASR calculated the mean μ_i and standard deviation σ_i of root mean square values of each PC across all 0.5-second windows and defined rejection threshold $T_i = \mu_i + k \cdot \sigma_i$ for each PC where k was a user-defined cutoff parameter. Later, each experimental data segment was again PCA decomposed and the threshold values derived from the reference data were used to determine whether the PCs would be kept or not. The PC components that were kept were back-projected to the sensor space, forming an artifact-removed version. Although the ASR method also uses magnitude as a criterion for removing artifacts, the cut-off threshold was different across PCs, which were dependent on the projection of the thresholds to the actual PC dimension. This means that a low-variance PC could

be discarded if the threshold projected to this PC dimension was too low. More discussion on this point will be provided in the Discussion section. The threshold parameter k used in the present application was 10 standard deviations, which is relatively conservative (*20*). The reference data used here was simply the artifact-free data cleaned by ICA. Note that this will reduce the rank of the data by the number of artifacts removed, resulting in the same number of zero variance PC in the subsequent PCA decomposition which did not cause technical issue in subsequent steps. Also, since the major signal variance is usually captured in the first few PCs, the reduced rank (causing zero variance PCs in the end) would not cause substantial effect. The two key parameters in ASR are: window length = 1 s; step size = 0.5 s.

Comparison of artifact handling methods. The comparison was conducted on two measures: computational cost and neural effect. In computational cost, we compared the three methods by reporting how much time (on average) they finish processing a single participant and a single trial in the visual oddball task. Each participant's data contains 160 trials of oddball stimulus, so the processing time for a single trial was simply calculated by the processing time for a participant divided by 160. The sampling rate of the data for calculating the computational cost is 150. The ICA algorithm used was from EEGLAB using the extended algorithm (Lee, et al., 1999). The computational costs for different stop criteria (sc) were reported. The computational cost calculated for ICA does not include the MARA automatic artifact detection. The key technical parameters of the computer are as follows: System: Windows 10 Enterprise; Processor: Intel(R) Core(TM) i7-8700 CPU @ 3.20 GHz; RAM: 16 G; Matlab version: R2020a. Two neural effects were compared: the early mismatch negativity (early effect) and the novelty-related P3 (late effect). The early neural effect was characterized as the ERP amplitude difference from 200 to 300 ms at O1 electrode for the visual oddball and at Cz electrode for the auditory oddball. The

late neural effect was characterized by ERP amplitude difference from 400 to 500 ms at Pz electrode for visual oddball and Cz electrode for auditory oddball. For testing whether the neural effects derived from the three methods are significantly different, we applied paired t-tests.

ERP-based neuroadaptive task for real-time experiment parameter optimization. After evaluating the SPA method in the performance of online artifact handling, we developed a real ERP-based neuroadaptive experiment and applied SPA to it. The main purpose of the experiment was to determine individually optimal inter-stimulus-interval (ISI) parameters that generated an ERP with the highest signal-to-noise ratio (SNR) in an online mode. Since ISI directly affects cognitive processes related to adaptation, learning, memory, mental effort and fatigue, which would in turn influence the generation of ERP, it is reasonable to assume that the optimal parameter differs across individuals. The task was a face processing task in which the participants were required to observe the facial pictures (from a public database: https://github.com/NVlabs/ffhq-dataset) presented serially and judge if the face was younger than 30 years old or not (by pressing a button for yes). The facial pictures were randomly presented in one square of a 2-by-2 lattice to avoid the adverse feeling when the faces are refreshed rapidly in a same location. No repetition of square position in the sequence was allowed. Each face stimulus stayed on the monitor until the next stimulus was shown. The duration of each face's presentation depended on the ISI which was manipulated (see below). The size of a single square in the 2-by-2 lattice was 540 pixels $\times$ 540 pixels. The screen resolution of the monitor was 1920 $\times$ 1080 pixels and the physical size was 52.6 $\times$ 29.5cm. The visual distance was about 60 cm to the monitor. Thus, the square forms a visual angle of 14 degrees. The EEG data were collected using mBrainTrain's Smarting amplifier (wireless) and was online broadcasted to Matlab via labstreaminglayer (LSL). The online EEG stream was

processed in Matlab and the ERP was extracted based on the presentation time of the face stimuli. SPA was used to remove the artifact and a Bayesian optimization module was applied to search the optimal ISI that leads to an ERP with the highest signal-to-noise ratio (SNR). Here, every segment from -300 ms to +700 ms after the presentation of the face image was sent to the SPA module for processing. Therefore, the online processing was actually with a temporal delay of more than 700 ms.

Bayesian optimization of ISI. During the serial presentation of facial pictures, the ISI was randomly changed every 10 seconds, ranging from 0.5 s to 5 s (corresponding to the 19 values of 10 seconds being divided by integers from 2 to 20). The Bayesian optimization approach was used to find individually optimal ISI that leads to the highest SNR of ERP averaged from a ten-second segment. The Bayesian optimization method has the advantage of finding the optimal parameter that lead to the minimum cost function with less samples (Mockus, 2012). The ERP derived from each 10-s block was obtained online and the SNR was defined and calculated as the ratio between the mean standard deviation (across electrodes) averaged from 200 ms to 500 ms and from the baseline period (-300 ms to 0 ms). Here, the sign-inversed SNR is the cost (objective) function to be minimized by Bayesian optimization, and the ISI is the parameter to optimize upon. Here, we used the Bayesian optimization implementation from the Matlab (R2020a) function '*bayesopt'*. Briefly, the Bayesian optimization first estimates the Gaussian process model (based on ARD Matérn 5/2 kernel, Snoek et al., 2012) of the objective function using the initial 4 samples of ISI as the burn-in phase for fitting the kernel. The kernel parameters to be estimated are length scale and signal standard deviation. The default initial values were used here. Specifically, the initial length scale was set to be the mean of standard deviations of the predictor (here, ISI), and the initial signal standard deviation was set to be the

standard deviation of the responses (here, SNR) divided by square root of 2. After the real-time updating phase starts, the probability of improvement acquisition function was used to identify the next ISI to be presented to the subject, and the Gaussian process model was updated incorporating newly acquired samples at each single iteration step.

Results

The bimodal distribution of PC amplitude from the raw EEG data

Figure 1 shows the amplitude distribution of all PCs from all single trials for the first 15 participants (Fig. 1A) and the result assembled from all 200 participants (Fig. 1B-C) from the visual oddball task. From the individual results, we can see that the bimodal distribution pattern exists in most of the participants. Around 90% of all individuals exhibit this bimodal distribution pattern, i.e., a clearly differentiable bimodality, indicating the dichotomy between large-variance artifacts and other components. The upper boundary of the small-variance cluster for the present dataset is at around 30 μV. In the grand average pattern (Fig. 1C), the upper peak in the bimodal distribution was blurred (Fig. 1C) due to considerable individual differences (Fig. 1B).

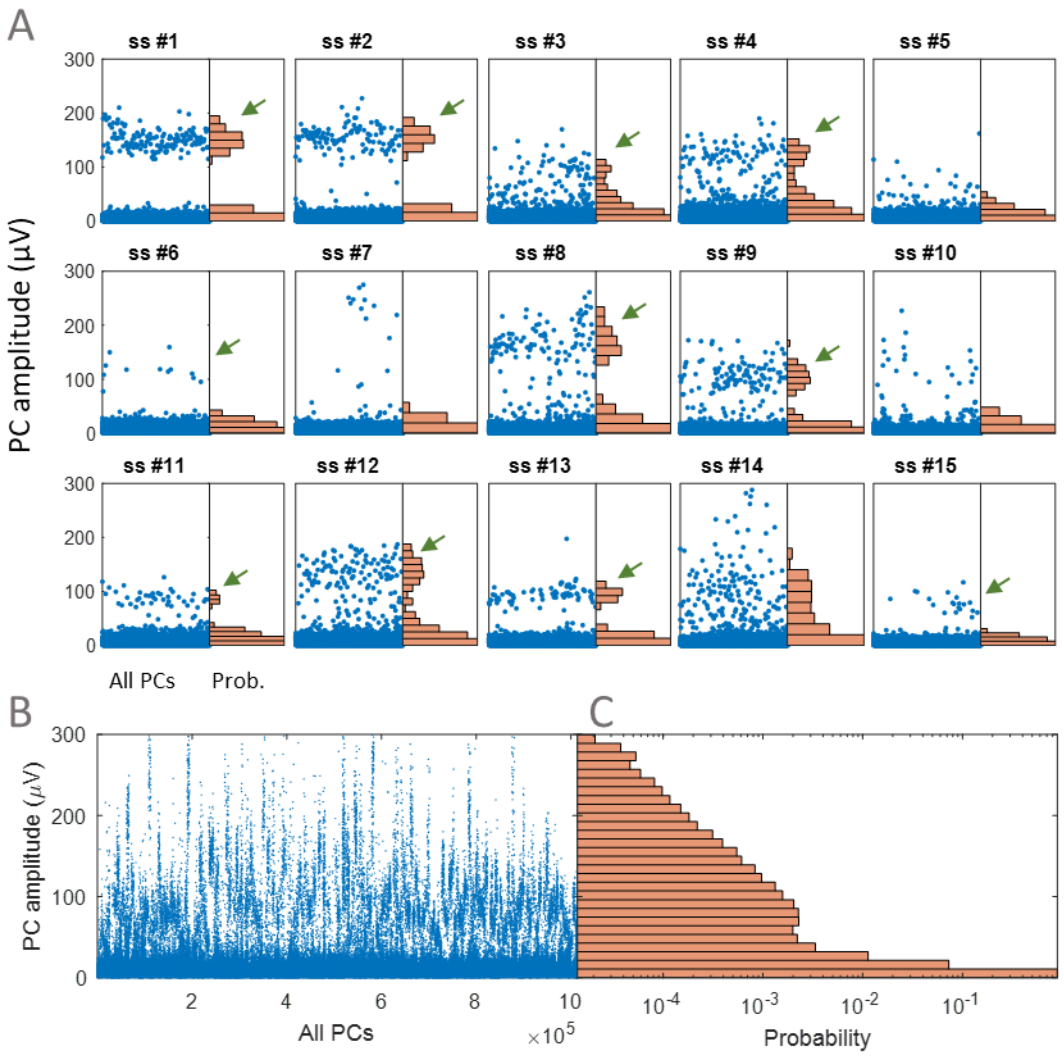

Figure 1. Bimodal distribution of PC amplitudes. **A.** The PC amplitude for all PCs from all single trials (left panel) and the probability distribution (right panel) from the first 15 participants from the visual oddball task. **B.** The results assembled from all participants.

To examine whether the bimodal patterns as shown in Figure 1 differentiate large-variance artifacts and other components, we calculated the distribution of PC amplitudes from the same dataset but after being cleaned by ICA and MARA. The new distribution patterns are plotted in Figure 2, which shows that the bimodal pattern of the distribution of PC amplitude is completely

gone after ICA-based artifact removal with the upper mode appearing to be truncated. This result suggests that these large-variance PCs were mostly identified as artifacts by ICA decomposition and the machine learning-based artifact identification algorithm MARA.

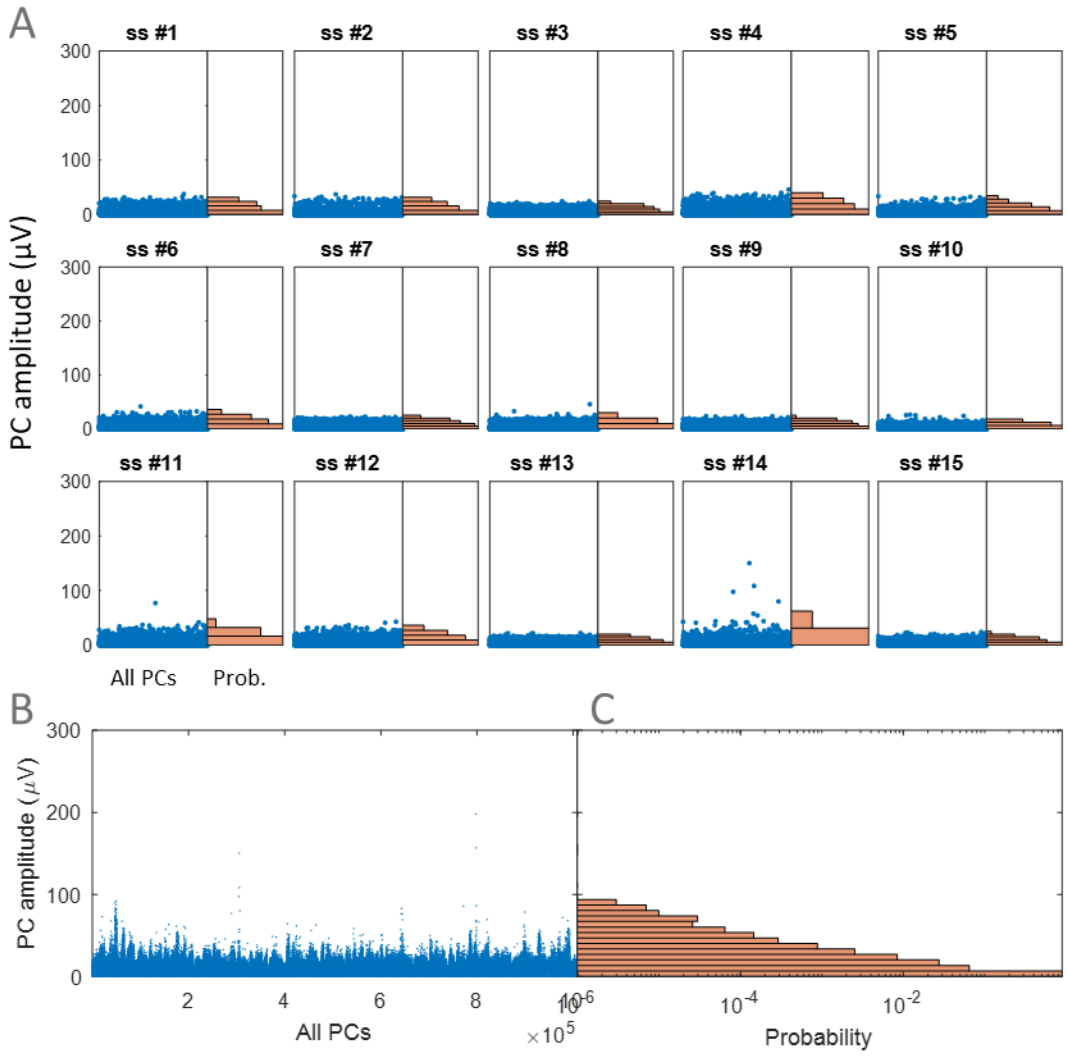

Figure 2. Distribution of PC amplitudes after ICA-based artifact removal. **A.** The PC amplitude for all PCs from all single trials (left panel) and the probability distribution (right panel) from the first 15 participants from the visual oddball task. **B.** The results assembled from all participants.

Removal of artifact using SPA

Based on the existence of the bimodality, a threshold can be determined that separates the PCs in different modes (clusters) in the distribution. For illustration purpose, Figure 3 shows the effect of removing the PCs in the upper peak in the variance distribution in a single participant. The threshold used is 30 (PC amplitude). Before the removal of PC components, clear artifacts (mainly ocular) can be seen in the single trials (Fig. 3B-C) which is confirmed by the scalp maps mainly covering frontal and forehead areas (Fig. 3B-C). After the removal of large-variance PCs in the upper distribution peak based on the threshold, the ERP patterns appear to be more neural-like as shown by the scalp maps in Fig. 3D-E.

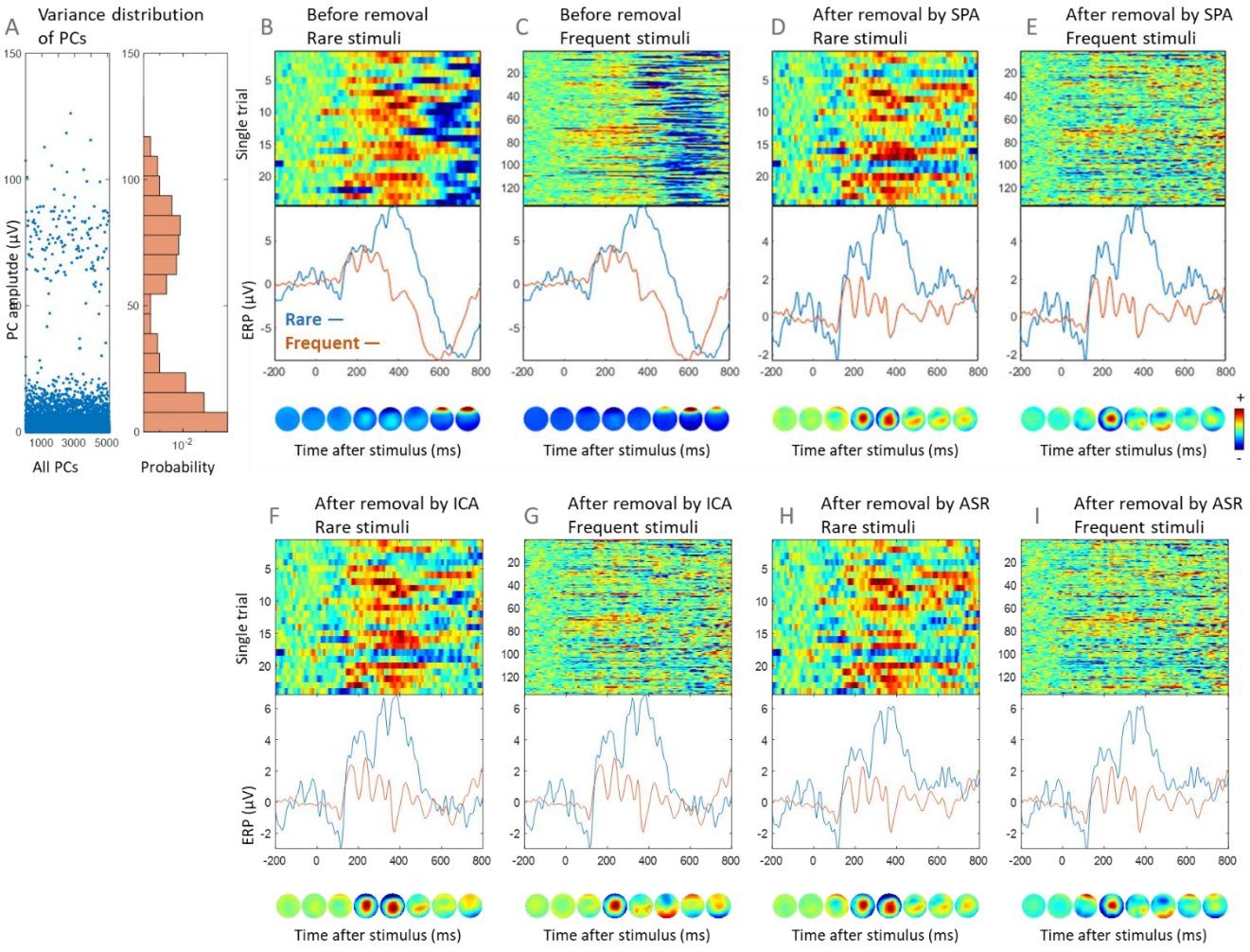

Figure 3. Removal of major PCs components with a distinct artifact feature. **A.** The distribution of PC amplitude of all PCs derived from all single trials. **B-D.** Single trials, average, and scalp map of ERP before (**B-C**) and after (**D-E**) the removal of large-variance PCs by SPA. The scalp maps correspond to the time points on the time axis they vertically align to. For comparison, we also show the results from ASR and ICA in F-I, which show a consistent effect.

The scalp map patterns of the artifact PCs remove by SPA

Based on the clustering algorithm, the average scalp maps of the first ten clusters are shown in Fig. 4 for both visual and auditory tasks. The results demonstrate that the artifacts removed are dominated by ocular movements and bad electrode connections.

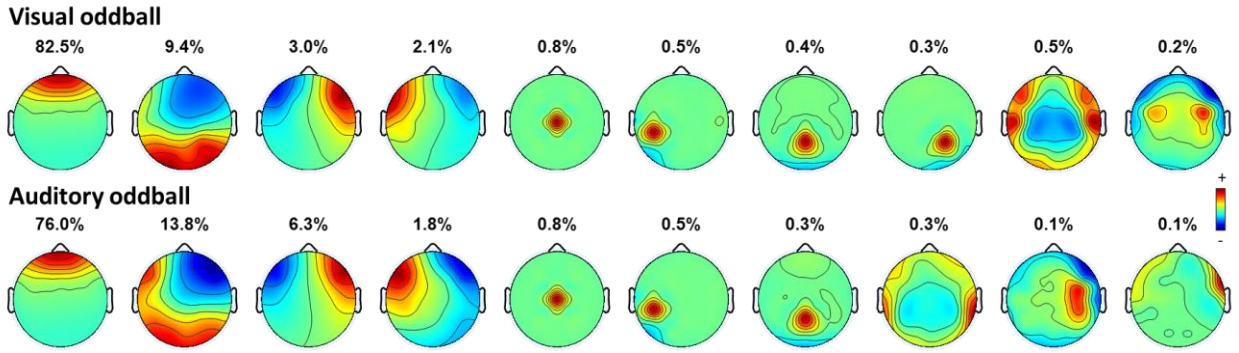

Figure 4. The scalp map patterns of the major PCs removed by SPA. Only the first 10 clusters were shown. The occupation percentages of different clusters (from the pool of all artifact PCs in all 200 participants) were marked above the scalp maps.

Comparison of SPA and other methods in real data

Figure 5 shows the ERP waveforms from the electrode Pz from the raw data and from the data processed by different methods. It is worth noting that the raw data have been bandpass filtered between 1 and 40 Hz, and average referenced. The results for both individuals and grand average

are shown. The raw data contains a considerable amount of artifact mostly due to ocular activity, which injects a large portion of variance across participants (Fig. 5A-B) and distorts the grand average ERP waveforms (Fig. 5C) as well as the between-condition difference (neural effect). The distortion of the oddball effect by the artifact can be clearly seen by the scalp map of the neural effect which shows a dominant ocular pattern (Fig. 5C). After ICA-based artifact correction, the artifacts were effectively removed and the neural effects were appropriately represented (Fig. 5F). Both the SPA and ASR methods appear to also effectively remove the artifact while preserving the major neural effects in a way that highly resembles ICA's results, indicating the feasibility of applying fast algorithms for online removal of major artifact while preserving the underlying neural effects for the use of neuroadaptive task paradigms. The computational cost for the three methods are summarized in Table 1. It is worth noting that the ICA algorithm used here has to be used in the entire data, thus reporting the average time used for a single trial does not actually make sense, nevertheless, it helps to provide a perspective on the computational cost. On average, the SPA algorithm only costs 0.8 millisecond for processing each single trial.

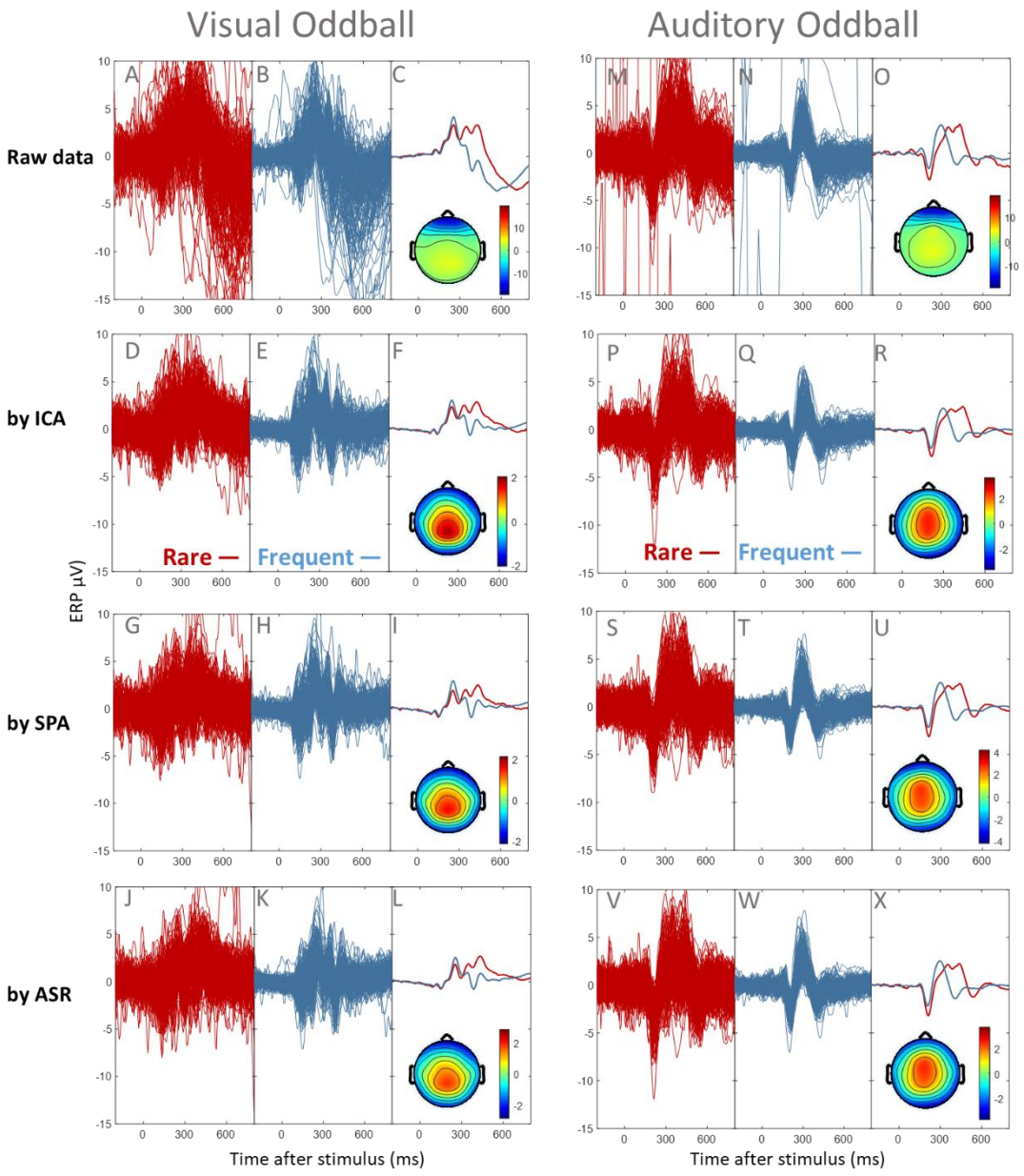

Figure 5. Removal of artifacts by different methods. **First column:** the ERP waveforms for rare stimuli from Pz for all participants. **Second column:** the ERP waveforms for frequent stimuli from Pz for all participants. **Third column:** The grand average ERPs waveforms for the two conditions. The scalp maps are from 400-500 ms. The first row is from the raw data (bandpass filtered and average-referenced). The second to last rows are for the results processed by different methods.

	ICA (sc = .001)	**ICA** (sc = .01)	**ICA** (sc = .1)	**ASR**	**SPA**
per participant (160 trials)	69.7 s	44.9 s	10.9 s	1030 ms	130 ms
per trial	436 ms	280 ms	68 ms	6.4 ms	0.8 ms

Table 1. The computational cost for different methods in removing the artifacts averaged across participants (sc: stopping criteria).

Comparison of the preserved neural effects across methods

We now examine how well SPA and ASR methods preserve the neural effects of early mismatch negativity and late novelty-related positivity in the ERP data (will be termed as early and late neural effects hereafter).

Descriptively, the purified ERP waveforms, neural effects, and scalps maps of the effects show a high degree of consistency across the three methods (Fig. 6 A-L). Statistically, the neural effects, in terms of mean level, are not differentiable between ICA and SPA in three out of the four examined effects, and are not differentiable between ICA and ASR in two out of the four examined effects (Fig. 6N-P). Although some systematic differences between ICA and the two fast algorithms exist, the general pattern and effect size are very close across the methods. The systematic differences are directly determined by various parameter settings of the three methods and EEG preprocessing such as thresholds, stopping criterion, filtering bands. Therefore, it is important to note that these statistical results are based on the current two hundred participants, under the current parameter setting, and they mainly serve to reflect the degree of similarity and difference among the methods but not to demonstrate the statistical distinguishability among them. Overall, the differences between the two fast algorithms and ICA in the magnitude of neural effects are relatively small. If we take the ICA-based artifact removal as a benchmark, the deviations of the two methods in terms of percentage are as follows. For SPA, early effect in

visual oddball: 5.6%, late effect in visual oddball: 4.0%, early effect in auditory oddball: 16.0%, late effect in auditory oddball: 3.8%. For ASR: early effect in visual oddball: 9.5%, late effect in visual oddball: 11.8%, early effect in auditory oddball: 20.0%, late effect in auditory oddball: 6.1%. A noticeable difference between ICA and the other two fast algorithms was found in the early effect of auditory oddball paradigm, which will be discussed later.

In terms of cross-individual variability, the effects obtained by the three methods showed to be highly correlated (Fig. 6Q-R) across participants (correlation coefficients *r* are shown alongside the scatter plots). No systematic bias was observed.

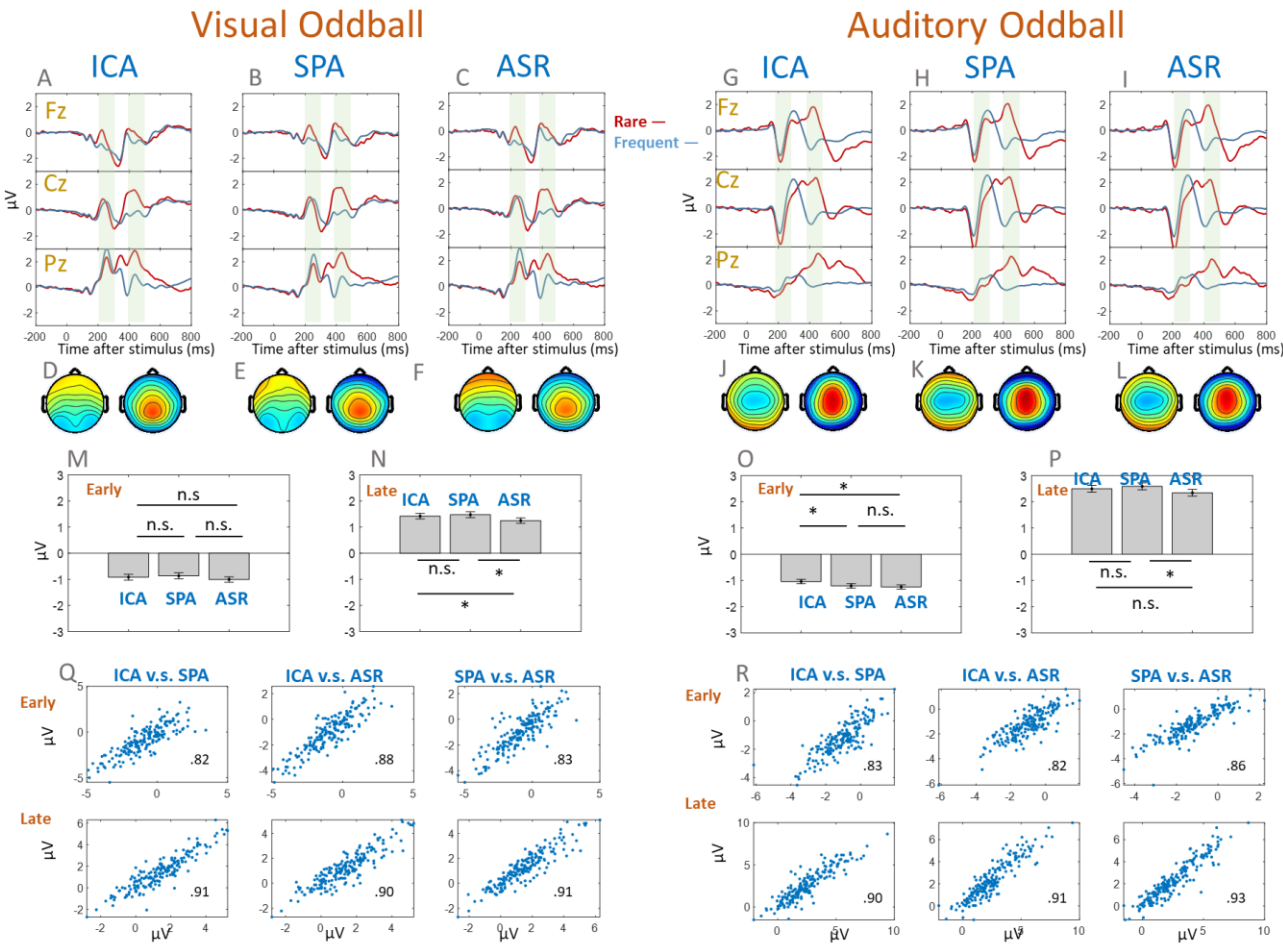

Figure 6. Comparison of neural effects across methods. **A-L:** The ERP waveforms and scalp maps of the neural effects. Light green bars on the ERP waveforms indicate the time windows used to generate the scalp maps of the neural effects (ERP differences). **M-P:** Statistical comparisons of the neural effects obtained by different methods. The star indicates that the difference is significant (n.s.: not significant). **Q-R:** The cross-subject correlation of neural effects obtained by different methods.

Online searching of individually optimal ISI based on Bayesian optimization

The comparison analysis showed that SPA generated a clean version of ERP close to the benchmark method while having a superior computational efficiency. In this section, we applied the SPA algorithm to a real neuroadaptive experiment that requires online acquisition of ERPs. Based on the set-up as illustrated by Figure 7A, the optimal ISI that led to the highest SNR of ERP for each individual participant was searched by Bayesian optimization. Figure 7B (left) shows an example of the searching process from the first participant. It has to be noted that although 120 points were sampled in the experiment, the optimal ISI for this participant had been reached in the first few points. The convergence process is certainly dependent on hyperparameters of the search algorithm. For comparison, Figure 7B (right) shows the result from surrogate ERP data without ISI dependence, from which the optimal solution cannot be found. The surrogate ERP data were simply generated by broadcasting an existing sample EEG dataset (instead of real EEG stream collect from a participant in real time) to the online optimization module. Figure 7C shows that this online searching scheme had successfully found optimal ISIs for all of the five participants. As expected, there are slight differences in the identified optimal ISI across participants (The optimal ISIs were shown by the red numbers in Figure 7D). Figure 7D shows that the searching process has arrived at points close to the

converged solution after sampling of certain amount of data, which also differs across participants, ranging from a few to dozens of points. Only 60 steps were shown here (total searching step was 120).

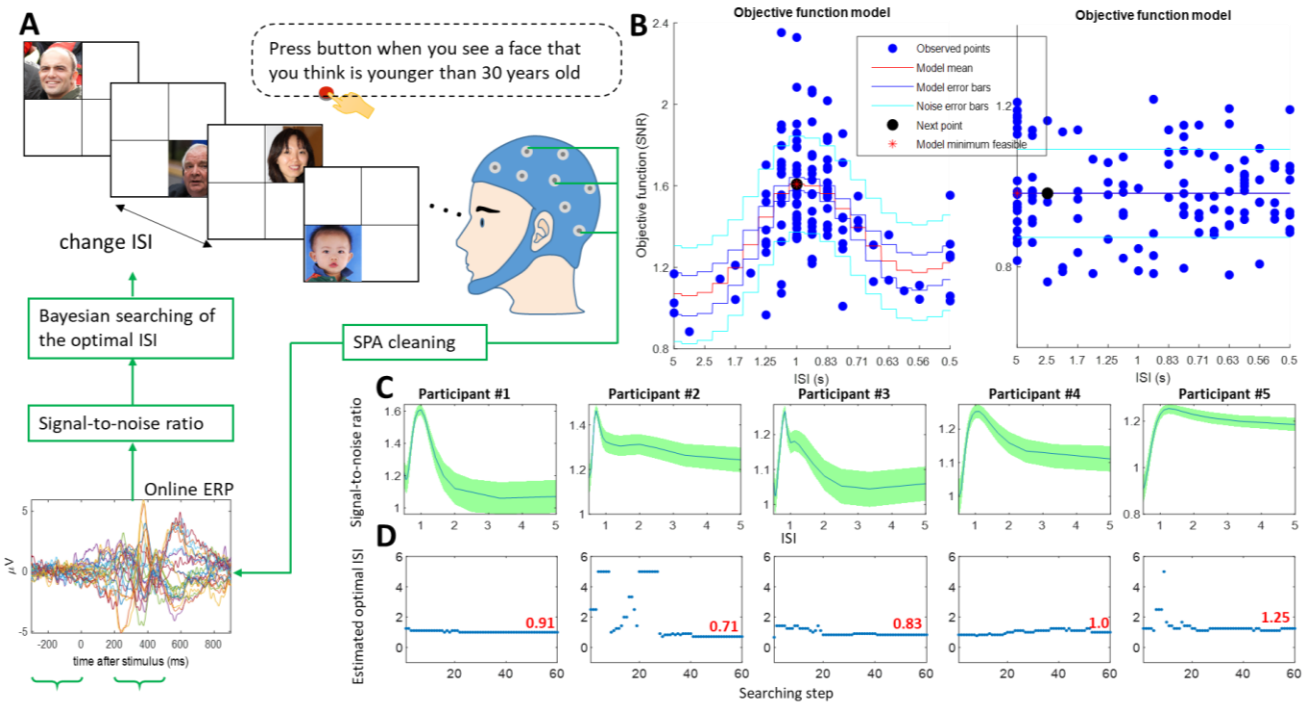

Figure 7 Bayesian optimization-based searching of individually optimal ISI that generates highest SNR of ERP. **A:** Illustration of the ERP-based neuroadaptive experimental paradigm. **B:** Examples of the searching process from participant #1 (left) and from surrogate ERP data without ISI dependence (right). **C:** The estimated objective functions for all participants after the searching. The line and shaded area represent the mean and uncertainty of the Gaussian process model. **D:** The estimated optimal ISI as a function of search steps.

Discussion

In this article, we developed the major modules and investigated the feasibility of ERP-based neuroadaptive research paradigm in which the online ERP is obtained and individually optimal experimental parameters are identified in real time. As a major component for this system, we first presented a simple algorithm SPA for rapidly removing EEG artifacts in an online mode and evaluated its performance in preserving relevant neural activation components and effects. By using SPA as the module for online artifact removal, we further developed and demonstrated an ERP-based neuroadaptive paradigm that applies Bayesian optimization method to search individually optimal experimental parameter online. This study is the first time that demonstrates the feasibility of using ERP for neuroadaptive research.

With its low computational cost and high performance in preserving ERP pattern and effects, the SPA method appears to be a suitable method for online handling of artifacts for instantaneous acquisition of brain response pattern as characterized by ERP. Nevertheless, there are several open questions, caveats and limitations that still remain, which is elucidated below.

SPA cannot replace advanced offline methods that isolates artifacts based on high order statistical features. Although the current SPA results showed good performance in preserving relevant neural effects in the oddball tasks, it does not mean that it should be used in offline analysis for artifact removal. It is solely intended to provide fast rough artifact correction for neuroadaptive paradigms where more accurate offline procedures are not feasible. Actually, it is not necessary and not recommended to use SPA in offline EEG preprocessing for a conventional basic neurocognitive study in which computational cost is not of major concern but precision and standardization are (Desjardins et al., 2021). Moreover, even in an online application, SPA can be applied along with other procedures that aim to more strictly clean the data, including online

problematic electrode detection, problematic temporal segment detection, context specific band-pass filtering, etc.

Comparison of SPA and ASR. Regarding the differences between the two fast algorithms SPA and ASR, there are three: 1) ASR requires clean reference data to calculate the mean and variance of activity in each PC axis, which is used to determine the threshold for removing artifactual components in the online data. 2) ASR treats each PC dimension (obtained from the online data segment) separately, i.e., the thresholds for each dimension are different. The threshold in each PC dimension is determined by the summed projection of the thresholds from all PC dimensions from the calibration data. This means that some PC components could be removed even when they do not contribute large variance to the signal at sensor level because the projected threshold to those PCs are small. This is a major difference from SPA because, in SPA, the classification of artifacts is directly related to variance contributed. This may explain why ASR appears to attenuate the neural effects (Fig. 6N-P) even when a relatively conservative threshold is used. 3) Computationally, SPA is much less costly than ASR. ASR does have an advantage as compared to SPA when the feature to be obtained is not ERP averaged from single trials, but the neural features of continuous signals (e.g., spectral features, commonly used in BCI). In this case, ASR is more suitable because it is designed to process continuous EEG signals with a smoothing operation to smooth the connection between adjacent data segments.

Determination of the threshold. A bimodal distribution pattern of PC amplitudes was found from the majority of participants in the present data, thus supporting the application of a simple threshold-based algorithm as adopted by SPA. The PCs in the high amplitude cluster almost completely disappear after a standard ICA-based artifact removal procedure (Fig. 2), which supports their artifactual nature. The threshold (upper boundary of the small variance cluster)

appears to be at around 30 μV for the present data. However, due to potential variations from lab to lab, it is recommended that the threshold should be determined based on the bimodality pattern from the actual data in the users' lab. The threshold may even be determined on an individual basis, which would incur more time and effort in preparation but not in online applications.

Reference issue. Because PCA extracts its first component with the largest variance, it should be best applied on average-referenced data. In EEG data that is not average-referenced, EEG activity that occurs at the reference site will be magnified as it is redistributed to the rest of the EEG channels, making it more likely that it will be captured by large-variance PCs and be discarded.

Demonstration of the feasibility of using ERP as a neural feedback for optimizing individual-specific experimental parameters in real time. Conventionally, ERP studies are designed based on pre-determined parameters and the results from the entire participant cohort are analyzed after data collection. This nature makes it infeasible to systematically evaluate the suitability of the experimental parameters, globally or individually. This limitation is one of the reasons that sparked the proposals of closed-loop neuroadaptive paradigms that aim to search optimal parameters online (Chen & Pesaran, 2021; Lorenz, et al., 2017). However, it is still unclear if online obtained ERP from a relatively short time period is able to provide sufficiently accurate information to drive a neuroadaptive study in a practical way or not. One concern is that the ERP effects are usually too subtle to be detected online, especially when the ERP is obtained from a short time session. If the variation of ERP (associated with a factor of interest) is too weak, it may not be able to effectively drive an optimization algorithm as the samples of the

objective functions may be too ‘noisy’, that is, not able to converge to a clearly structured pattern from which an optimum is to be detected.

We applied the SPA online artifact handling module and, for the first time, demonstrated that online acquisition of in-session ERP can serve to establish a neuroadaptive paradigm in which individually optimal experimental parameter is determined online. For this proof-of-concept demonstration, the parameter to be optimized here was inter-stimulus-interval (ISI) and the objective function was signal-to-noise ratio (SNR) of ERP obtained from a 10-second session. Conventionally, the ISI parameter is usually heuristically determined in an ERP study. The reason of choosing SNR is that it creates an inverse U-shaped relationship (see reasons below) in which the process of Bayesian-based optimization is better visualized, whereas other metrics (e.g., ERP amplitude or effect) may not easily show an inverse U-shaped relationship. We hypothesized the existence of an individually specific ISI that leads to the largest SNR based on the following simple rationales: 1) too long ISIs lead to fewer trials within a unit amount of time, thus reducing the SNR of ERP; 2) too short ISIs overload the cognitive resource, thus diminishing ERP amplitude and consequently reducing the SNR of ERP. Our application of Bayesian optimization successfully identified individually optimal ISIs in all five participants. The searching process converged to the solution within only a few minutes. This application demonstrated the feasibility of using online obtained ERP as a neural feedback for optimizing experiment parameters. We want to emphasize, that from our results, we do not infer that other methods are not capable of implementing the online optimization, nor do we want to convey that SPA achieved the best result. Our study rather serves to demonstrate the ability of SPA in supporting online optimization with a superior computational efficiency. In addition, it should be noted that the optimal ISIs that we identified for each individual in the present study cannot yet

to be claimed as “ground truth”. What was demonstrated in the present study is that an optimal value cannot be obtained in surrogate ERP data where ERP’s dependence on ISI did not exist (Figure 7B, right). To tackle the ground truth issue, we would need to conduct multi-session experiments in the same participant and examine the test-retest reliability in the future.

It is worth noting that our implementation of an ERP-based neuroadaptive paradigm utilizes in-session (i.e. within-block) trial-averaging to achieve a better representation of neural response patterns to provide online neural feedback at a time scale of seconds. The trial-averaging has the advantage of cancelling out spontaneous neural activity that is usually not recognized as artifacts by artifact handling methods but still affect the neural response pattern (ERP) concerned. Therefore, in principle, averaging over a few trials improves neural response representation but lengthens the experiment duration, which forms an inherent trade-off. This principle would apply to any online artifact handling methods even if they are able to handle single trial-based feedback. The performance of handling single trials as well as determining the optimal number of trials to average for neuroadaptive study would heavily depend on individual studies and their respective contrast-to-noise ratio.

Contribution to the generalizability issue in ERP research. The ability to identify latent parameters seated in a complex, individual neural system that are relevant to a neural cognitive process of interest may improve the generalizability of research findings about the process. Individual brains substantially differ from each other in various aspects. In neuroimaging research, a common strategy is to use a localizer task to adjust for individual variability in the location of a functional area such as the visual word form area (e.g., Glezer & Riesenhuber, 2013). In this study, we showed that individually optimal ISIs under certain objective function (here, SNR of ERP) can be obtained in real time. Importantly, the Gaussian process explicitly

models a subject's brain response across many different ISIs (i.e., we are not just obtaining a single optimal solution). This allows drawing inference about the range of ISIs for which certain ERP characteristics can be studied within an individual subject, or also across a cohort of subjects (for example by using hierarchical GPs to combine observations across subjects, Hensman et al., 2013). In turn, this can give us important insights about how generalizable an ERP finding is to a range of different ISIs, and equally informs us about the range of ISIs for which certain ERP characteristics break down (which in itself could be an interesting topic of research). Taking into consideration that the search space can be expanded by far more dimensions of interest such as different experimental settings, stimuli, procedures (for a discussion about this aspect, please see Yarkoni et al. 2019) or even other cognitive tasks (e.g., Lorenz et al. 2018, 2021), the approach shows potential in allowing for far more principled generalization of the results obtained.

Conclusion. The present work showed that the simple and extremely fast algorithm SPA based on removing PCs with distinctively large variance performs well in preserving ERP patterns and relevant neural effects. Our application of SPA in online extraction of ERPs also demonstrated the feasibility of ERP-based neuroadaptive paradigms for optimizing experimental parameters within the online data collection session. We hope that our findings motivate ERP researchers to explore the exciting potential of ERP-based neuroadaptive paradigms for systematically investigating the generalizability of ERP findings and leveraging it clinically to accelerate ERP-based biomarker discovery.

Acknowledgments

This work was partially supported by the Hong Kong Research Grant Council (ECS 27603818, GRF 17609321) and the Seed Fund for Basic Research from the University of Hong Kong (202011159042).

References

Afergan, D., Peck, E. M., Solovey, E. T., Jenkins, A., Hincks, S. W., Brown, E. T., ... & Jacob, R. J. (2014, April). Dynamic difficulty using brain metrics of workload. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3797-3806).

Andrade, A.O., et al., EMG signal filtering based on Empirical Mode Decomposition. Biomedical Signal Processing and Control, 2006. 1(1): p. 44-55.

Arns, M., Heinrich, H., & Strehl, U. (2014). Evaluation of neurofeedback in ADHD: the long and winding road. *Biological psychology*, *95*, 108-115.

Belouchrani, A., et al., A blind source separation technique using second-order statistics. Ieee Transactions on Signal Processing, 1997. 45(2): p. 434-444.

Berg, P. and M. Scherg, A Multiple Source Approach to the Correction of Eye Artifacts. Electroencephalography and Clinical Neurophysiology, 1994. 90(3): p. 229-241.

Bertrand, O., F. Perrin, and J. Pernier, A Theoretical Justification of the Average Reference in Topographic Evoked-Potential Studies. Electroencephalography and Clinical Neurophysiology, 1985. 62(6): p. 462-464.

Chen, Z. S., & Pesaran, B. (2021). Improving scalability in systems neuroscience. *Neuron*.

Chang, C.Y., et al., Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings. Ieee Transactions on Biomedical Engineering, 2020. 67(4): p. 1114-1121.

Daly, I., et al., Automated Artifact Removal From the Electroencephalogram: A Comparative Study. Clinical Eeg and Neuroscience, 2013. 44(4): p. 291-306.

Daly, I., et al., FORCe: Fully Online and Automated Artifact Removal for Brain-Computer Interfacing. Ieee Transactions on Neural Systems and Rehabilitation Engineering, 2015. 23(5): p. 725-736.

da Costa, P. F., Haartsen, R., Throm, E., Mason, L., Gui, A., Leech, R., & Jones, E. J. (2021). Neuroadaptive electroencephalography: a proof-of-principle study in infants. *arXiv preprint arXiv:2106.06029*.

de Freitas, A.M., et al., EEG artifact correction strategies for online trial-by-trial analysis. Journal of Neural Engineering, 2020. 17(1).

Delorme, A. and S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, 2004. 134(1): p. 9-21.

Delorme, A., T. Sejnowski, and S. Makeig, Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. Neuroimage, 2007. 34(4): p. 1443-1449.

Desjardins, J. A., van Noordt, S., Huberty, S., Segalowitz, S. J., & Elsabbagh, M. (2021). EEG Integrated Platform Lossless (EEG-IP-L) pre-processing pipeline for objective signal

quality assessment incorporating data annotation and blind source separation. *Journal of Neuroscience Methods*, *347*, 108961.
Dien, J., Evaluating two-step PCA of ERP data with Geomin, Infomax, Oblimin, Promax, and Varimax rotations. Psychophysiology, 2010. 47(1): p. 170-83.
Dien, J., Issues in the application of the average reference: Review, critiques, and recommendations. Behavior Research Methods Instruments & Computers, 1998. 30(1): p. 34-43.
DiMattina, C., & Zhang, K. (2013). Adaptive stimulus optimization for sensory systems neuroscience. *Frontiers in neural circuits*, *7*, 101.
Egambaram, A., et al., Automated and Online Eye Blink Artifact Removal from Electroencephalogram. Proceedings of the 2019 Ieee International Conference on Signal and Image Processing Applications (Ieee Icsipa 2019), 2019: p. 159-163.
Glezer, L. S., & Riesenhuber, M. (2013). Individual variability in location impacts orthographic selectivity in the "visual word form area". J Neurosci, 33(27), 11221–11226.
Hensman, J., Lawrence, N. D., & Rattray, M. (2013). Hierarchical Bayesian modelling of gene expression time series across irregularly sampled replicates and clusters. BMC bioinformatics, 14(1), 1-12.
Hassanien, A. E., & Azar, A. A. (2015). Brain-computer interfaces. Switzerland: Springer.
Hoedlmoser, K., Pecherstorfer, T., Gruber, G., Anderer, P., Doppelmayr, M., Klimesch, W., & Schabus, M. (2008). Instrumental conditioning of human sensorimotor rhythm (12-15 Hz) and its impact on sleep as well as declarative learning. *Sleep*, *31*(10), 1401-1408.
Jiang, X., G.B. Bian, and Z.A. Tian, Removal of Artifacts from EEG Signals: A Review. Sensors, 2019. 19(5).
Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., Mckeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. Psychophysiology, 37(2), 163-178.
Kangassalo, L., M. Spape, and T. Ruotsalo, Neuroadaptive modelling for generating images matching perceptual categories. Scientific Reports, 2020. 10(1).
Kobler, R.J., et al., Corneo-retinal-dipole and eyelid-related eye artifacts can be corrected offline and online in electroencephalographic and magnetoencephalographic signals. Neuroimage, 2020. 218.
Kothe, C. A. E., & Jung, T. P. (2016). *U.S. Patent Application No. 14/895,440*.
Krol, L. R., Andreessen, L. M., & Zander, T. O. (2018). *Passive brain–computer interfaces: a perspective on increased interactivity* (pp. 69-86). CRC Press.
Krol, L. R., Haselager, P., & Zander, T. O. (2020). Cognitive and affective probing: a tutorial and review of active learning for neuroadaptive technology. *Journal of neural engineering*, *17*(1), 012001.
Lawhern, V., et al., Detection and classification of subject-generated artifacts in EEG signals using autoregressive models. Journal of neuroscience methods, 2012. 208(2): p. 181-189.
Lee, T.W., M. Girolami, and T.J. Sejnowski, Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. Neural Computation, 1999. 11(2): p. 417-441.
Lorenz, R., A. Hampshire, and R. Leech, Neuroadaptive Bayesian Optimization and Hypothesis Testing. Trends in Cognitive Sciences, 2017. 21(3): p. 155-167.

Lorenz, R., Simmons, L. E., Monti, R. P., Arthur, J. L., Limal, S., Laakso, I., ... & Violante, I. R. (2019). Efficiently searching through large tACS parameter spaces using closed-loop Bayesian optimization. *Brain stimulation*, *12*(6), 1484-1489.

Lorenz, R., Violante, I. R., Monti, R. P., Montana, G., Hampshire, A., & Leech, R. (2018). Dissociating frontoparietal brain networks with neuroadaptive Bayesian optimization. *Nature communications*, *9*(1), 1-14.

Lorenz, R., Monti, R. P., Violante, I. R., Anagnostopoulos, C., Faisal, A. A., Montana, G., & Leech, R. (2016). The Automatic Neuroscientist: A framework for optimizing experimental design with closed-loop real-time fMRI. *NeuroImage*, *129*, 320-334.

Lotte, F., Jeunet, C., Mladenović, J., N'Kaoua, B., & Pillette, L. (2018a). A BCI challenge for the signal processing community: considering the user in the loop.

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018b). A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of neural engineering*, *15*(3), 031005.

Roc, A., Pillette, L., Mladenovic, J., Benaroch, C., N'Kaoua, B., Jeunet, C., & Lotte, F. (2020). A review of user training methods in brain computer interfaces based on mental tasks. Journal of Neural Engineering.

Mannan, M.M.N., M.A. Kamran, and M.Y. Jeong, Identification and Removal of Physiological Artifacts From Electroencephalogram Signals: A Review. Ieee Access, 2018. 6: p. 30630-30652.

Mockus, J. (2012). Bayesian approach to global optimization: theory and applications (Vol. 37). Springer Science & Business Media.

Morlet, D., & Fischer, C. (2014). MMN and novelty P3 in coma and other altered states of consciousness: a review. *Brain topography*, *27*(4), 467-479.

Mullen, T.R., et al., Real-Time Neuroimaging and Cognitive Monitoring Using Wearable Dry EEG. Ieee Transactions on Biomedical Engineering, 2015. 62(11): p. 2553-2567.

Näätänen, R., Kujala, T., Escera, C., Baldeweg, T., Kreegipuu, K., Carlson, S., & Ponton, C. (2012). The mismatch negativity (MMN)–a unique window to disturbed central auditory processing in ageing and different clinical conditions. *Clinical Neurophysiology*, *123*(3), 424-458.

Näätänen, R., Kujala, T., & Light, G. (2019). *Mismatch negativity: a window to the brain*. Oxford University Press.

Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. NeuroImage, 222, 117254.

Nguyen, H.A.T., et al., EOG artifact removal using a wavelet neural network. Neurocomputing, 2012. 97: p. 374-389.

Park, H.D. and O. Blanke, Heartbeat-evoked cortical responses: Underlying mechanisms, functional roles, and methodological considerations. Neuroimage, 2019. 197: p. 502-511.

Pei, L., & Ouyang, G. (2021). Online recognition of handwritten characters from scalp-recorded brain activities during handwriting. *Journal of Neural Engineering*, *18*(4), 046070.

Pion-Tonachini, L., K. Kreutz-Delgado, and S. Makeig, ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. Neuroimage, 2019. 198: p. 181-197.

Plochl, M., J.P. Ossandon, and P. Konig, Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. Frontiers in Human Neuroscience, 2012. 6.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128-2148.
Settles, B. (2009). Active learning literature survey.
Somers, B., & Bertrand, A. (2016). Removal of eye blink artifacts in wireless EEG sensor networks using reduced-bandwidth canonical correlation analysis. *Journal of neural engineering*, *13*(6), 066008.
Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, *25*.
Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
Urigüen, J. A., & Garcia-Zapirain, B. (2015). EEG artifact removal—state-of-the-art and guidelines. *Journal of neural engineering*, *12*(3), 031001.
Vidal, J. J. (1977). Real-time detection of brain events in EEG. *Proceedings of the IEEE*, *65*(5), 633-641.
Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & psychophysics*, *33*(2), 113-120.
Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M., & Shenoy, K. V. (2021). High-performance brain-to-text communication via handwriting. *Nature*, *593*(7858), 249-254.
Winkler, I., et al., Robust artifactual independent component classification for BCI practitioners. J Neural Eng, 2014. 11(3): p. 035013.
Winkler, I., S. Haufe, and M. Tangermann, Automatic classification of artifactual ICA-components for artifact removal in EEG signals. Behav Brain Funct, 2011. 7: p. 30.
Yarkoni, T. (2019). The generalizability crisis. *Behavioral and Brain Sciences*, 1-37.
Zander, T. O., Krol, L. R., Birbaumer, N. P., & Gramann, K. (2016). Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity. *Proceedings of the National Academy of Sciences*, *113*(52), 14898-14903.