

# Suppression of Cortical Electrostimulation Artifacts using Pre-whitening and Null Projection

J Lim<sup>1</sup>, P T Wang<sup>1</sup>, L Bashford<sup>2</sup>, S Kellis<sup>2</sup>, S J Shaw<sup>3</sup>, H Gong<sup>3</sup>, M Armacost<sup>3</sup>, P Heydari<sup>4,1</sup>, A H Do<sup>5</sup>, R A Andersen<sup>2</sup>, C Y Liu<sup>3</sup>, Z Nenadic<sup>1,4</sup>

1. Department of Biomedical Engineering, University of California Irvine (UCI), Irvine, CA 92697, USA

2. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA

3. Rancho Los Amigos National Rehabilitation Center, Downey, CA 90242, USA

4. Department of Electrical Engineering and Computer Science, UCI, Irvine, CA 92697, USA

5. Department of Neurology, UCI, Irvine, CA 92697, USA

E-mail: limj4@uci.edu

## Abstract.

*Objective:* Invasive brain-computer interfaces (BCIs) have shown promise in restoring motor function to those paralyzed by neurological injuries. These systems also have the ability to restore sensation via cortical electrostimulation. Cortical stimulation produces strong artifacts that can obscure neural signals or saturate recording amplifiers. While front-end hardware techniques can alleviate this problem, residual artifacts generally persist and must be suppressed by back-end methods.

*Approach:* We have developed a technique based on pre-whitening and null projection (PWNP) and tested its ability to suppress stimulation artifacts in electroencephalogram (EEG), electrocorticogram (ECoG) and microelectrode array (MEA) signals from five human subjects.

*Main results:* In EEG signals contaminated by narrow-band stimulation artifacts, the PWNP method achieved average artifact suppression between 32 and 34 dB, as measured by an increase in signal-to-interference ratio. In ECoG and MEA signals contaminated by broadband stimulation artifacts, our method suppressed artifacts by 78-80% and 85%, respectively, as measured by a reduction in interference index. When compared to independent component analysis, which is considered the state-of-the-art technique for artifact suppression, our method achieved superior results, while being significantly easier to implement.

*Significance:* PWNP can potentially act as an efficient method of artifact suppression to enable simultaneous stimulation and recording in bi-directional brain-computer interfaces to biomimetically restore motor function.

*Keywords:* brain-computer interface, stimulation artifacts, cortical stimulation, intracortical microstimulation, artifact suppression, electroencephalography, electrocorticography, intracortical microelectrode array. Submitted to: *J. Neural Eng.*

## 1. Introduction

Brain-computer interfaces (BCIs) are emerging as a promising solution for restoring communication to individuals with amyotrophic lateral sclerosis [1, 2, 3], or motor function to those with paraplegia [4, 5, 6] or tetraplegia [7, 8, 9]. Invasive BCIs based on electrocorticography (ECoG) [10, 11] or intracortical microelectrode arrays (MEAs) [12, 13] have the ability to elicit somatosensation via electrostimulation of cortical tissue. This makes it possible for BCIs, which currently primarily rely on visual feedback [4, 5, 6, 7, 8, 9], to enact closed-loop control using artificial somatosensory feedback in what is referred to as a “bi-directional” brain-computer interface (BD-BCI) [14]. Preliminary studies with BD-BCIs suggest that the artificial somatosensory feedback may improve BCI performance [15]. However, BD-BCIs, especially in motor applications, necessitate simultaneous stimulation and recording to enable smooth and continuous control. One critical requirement for this is that neural feature extraction must be performed in the presence of strong electrical stimulation artifacts. These artifacts can compromise BCI performance by masking neural activity or saturating front-end amplifiers [16, 17, 18]. To avoid this problem, BD-BCIs must be able to mitigate stimulation artifacts.

A necessary condition for artifact suppression strategies is that large-amplitude stimulation artifacts must not saturate analog front-ends. To this end, hardware innovations have been devised to increase front-end dynamic range [19, 20]. Other front-end approaches safeguard against saturation by employing a variety of techniques. Examples include template subtraction [21], adaptive filtering [22], and dipole cancellation [23, 24]. Even with the implementation of such methods, residual artifacts will persist and must be additionally suppressed by digital back-end methods.

The simplest of these back-end methods include blanking-reconstruction techniques, wherein data samples containing artifacts are removed and subsequently replaced by sample-and-hold [25, 26] or interpolated [27, 28, 29] data points. Another approach is to construct an artifact template by averaging artifact waveforms and subtract the template from the signal [30, 31]. However, these methods can create significant signal distortions due to data removal or replacement. A more elegant approach relies on signal decomposition techniques, such as independent component analysis (ICA) [32] or empirical mode decomposition (EMD) [33]. These techniques have been useful in separating artifact from neural sources [34, 35, 36, 37] and are thus considered state-of-the-art artifact suppression techniques. However, ICA-based methods may fail to separate stimulation artifacts and neural signals, as they are not guaranteed to be independent. Our recent study shows that this can happen even when stimulation artifacts are narrow-band [38]. Similarly, intrinsic mode functions generated by EMD [33] may fail to separate neural signals and stimulation artifacts because their local signal characteristics, including smoothness, may be similar. Finally, EMD’s numerical nature makes it suitable for offline analyses but not amenable to a real-time implementation.

Motivated by these shortcomings, we recently developed a subspace-based technique [38] that utilizes pre-whitening and null projection (PWNP) to efficiently separate artifact and neural signal subspaces. The artifacts can then be suppressed by projecting the contaminated signals away from the artifact subspace. This is in contrast to ICA, which requires intensive numerical optimization to decompose signals and a combinatorial search to identify the artifact components. When tested in a single subject whose electroencephalogram (EEG) was contaminated with electrical artifacts, our technique achieved superior suppression results to ICA, while having a much simpler implementation [38].

In this study, we demonstrate the efficacy of the PWNP method in suppressing cortical electrostimulation artifacts in a variety of neural data. Firstly, we expanded our preliminary analysis [38] to multiple subjects whose EEG data were contaminated by voltage artifacts introduced by a signal generator via scalp electrodes. Our analysis shows that PWNP can suppress these narrow-band artifacts, i.e., increase the signal-to-interference ratio (SIR), while preserving neuromodulation features, expressed by the signal-to-noise ratio (SNR). Secondly, we also analyzed subdural ECoG data collected in an epilepsy monitoring unit to demonstrate PWNP’s ability to suppress broad-band artifacts generated during cortical mapping procedures. Finally, we analyzed human MEA data collected during intracortical microstimulation (ICMS) to elicit artificial sensation. We show that PWNP can suppress these broad-band artifacts while preserving neural signals such as action potentials. Generally, the PWNP method outperformed ICA on these diverse datasets, while having a much simpler implementation. Therefore, we demonstrate that the PWNP method is an effective technique for the suppression of stimulation artifacts with obvious applications to BD-BCI technologies.

## **2. Methods**

### *2.1. Electrophysiological Data Collection and Pre-processing Procedures*

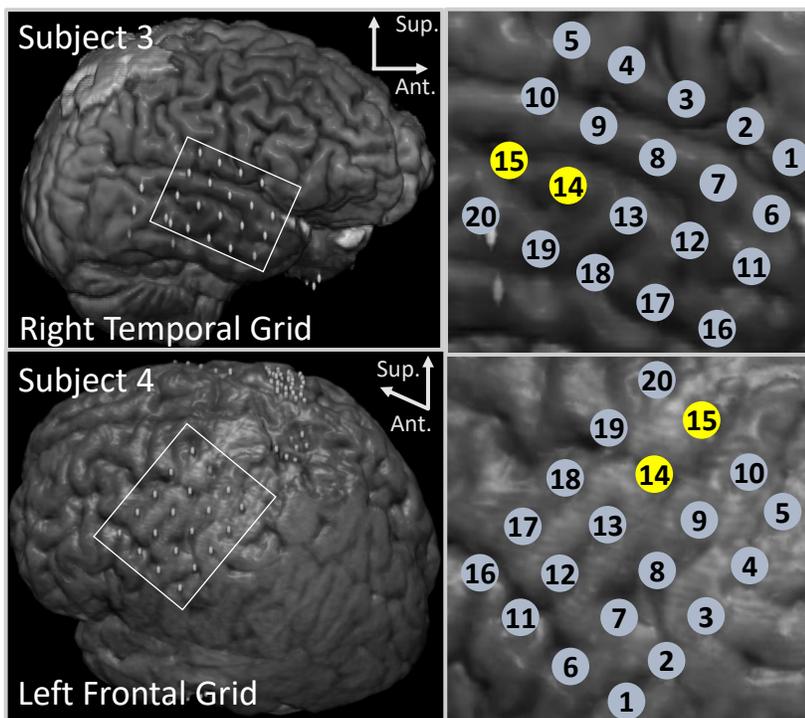
Data were collected with the informed consent of all subjects, and all procedures performed were approved by the Institutional Review Board of the University of California, Irvine, the University of Southern California, and the Rancho Los Amigos National Rehabilitation Center.

*2.1.1. EEG Data Collection* We collected data from two healthy volunteers (Subjects 1 and 2) using 20-electrode (10-20 international system), EEG caps (Compumedics USA, Charlotte, NC). We reduced the 30-Hz impedance of the electrode-scalp interface below 10 k $\Omega$  by applying conductive gel and abrading the scalp. Nineteen single-channel amplifiers (EEG100C, Biopac Systems, Goleta, CA) captured EEG signals with respect to a reference electrode located over the frontal lobe between Fp1/Fp2 and Fz electrodes. Amplifiers gains were set to 5000 $\times$  and band-pass filters were set to 1-35 Hz. These

signals were then sampled at 4000 Hz and digitized with a 16-bit acquisition system (MP150, Biopac Systems, Goleta, CA). The acquisition system also simultaneously recorded analog signals generated by a custom MATLAB script (Mathworks, Natick, MA) that tracked the behavioral cues displayed during the experiment.

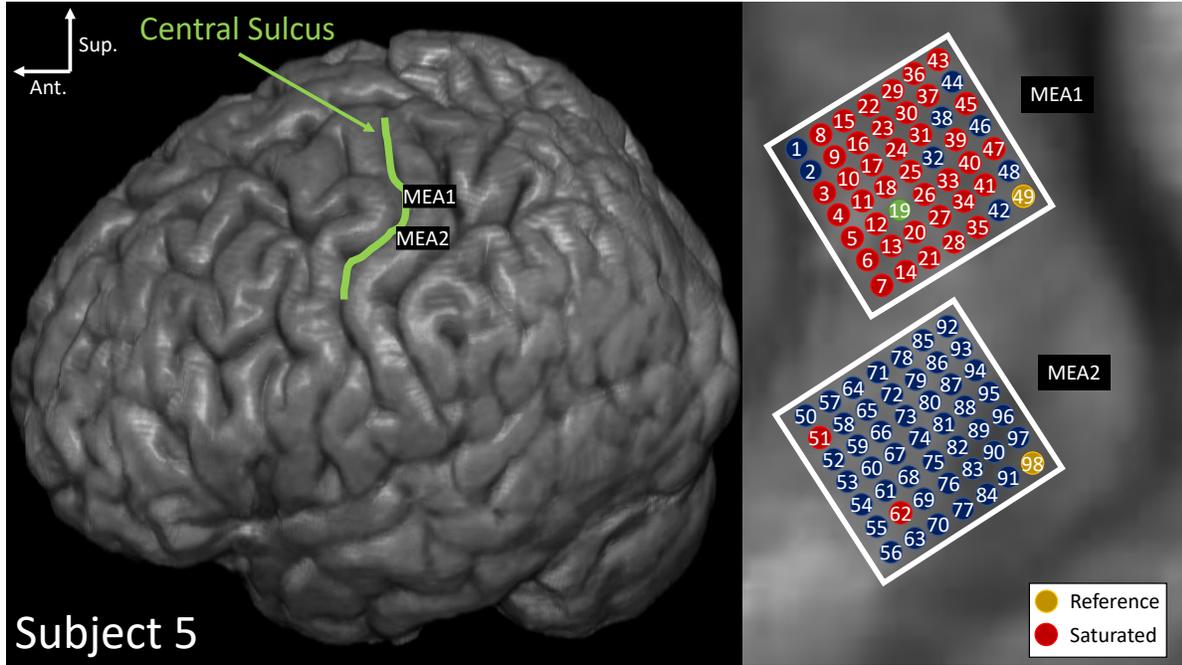
Since non-invasive cortical stimulation methods, like transcranial electrostimulation [39], are not commonly used in healthy volunteers, we utilized a hand-held, battery powered impedance meter (EIM105, General Devices, Ridgefield, NJ) as a surrogate for a cortical stimulator. Prior to placing the EEG caps, two individual EEG electrodes were affixed to the scalp with adhesive cream (EC2, Natus Neurology, Middleton, WI) to the left and right posterior of the Cz electrode. The impedance meter output, which was a 30-Hz voltage sine wave, was then introduced to the scalp via these electrodes in a bipolar configuration. This output was also recorded in parallel by the MP150 system. We first collected baseline activity for one minute without turning on the stimulation. Subsequently, subjects initiated the behavioral task, wherein the stimulator was turned on and subjects followed auditory cues that alternated between “eyes open” and “eyes closed”. Each eyes-open or eyes-closed epoch lasted 15 seconds, for a total of 20 epochs (5 minutes total). We then saved the EEG data for later analysis.

*2.1.2. ECoG Data Collection* We collected ECoG data at the hospital bedside from two subjects (Subjects 3 and 4) undergoing cortical electrostimulation as part of epilepsy surgery evaluation. Subject 3 was implanted with a standard size (10 mm spacing, 2.3 mm disc electrode diameter, platinum-iridium contacts) 4×5 ECoG grid (Integra Life Sciences, Plainsboro NJ) over the right temporal lobe. Subject 4 was implanted with the same type of grid over the left frontal cortex (see Figure 1). A clinical-grade bioamplifier (Natus<sup>®</sup> Quantum<sup>™</sup>, Natus Medical Incorporated, Pleasanton, CA) recorded ECoG signals at a sampling rate of 512 Hz. As part of eloquent cortex mapping procedures, an FDA-approved cortical stimulator (Nicolet<sup>®</sup> Cortical Stimulator, Natus Medical Incorporated, Pleasanton CA) delivered square pulse trains across a pair of electrodes (stimulation channel). Each pulse train was delivered for a predetermined duration (stimulation epoch) across a range of current amplitudes (2-10 mA). Both subjects were stimulated with the following parameters: 50 Hz pulse train frequency, 250  $\mu$ s pulse width, 5 second stimulation epoch, and current amplitudes ranging from 2 to 10 mA in increments of 2 mA. We saved the collected ECoG data for later analysis in MATLAB. Data from the electrodes comprising the stimulation channel were excluded as they saturated during stimulation. All data were high-pass filtered (4<sup>th</sup> order, Butterworth, 1.5 Hz, zero-phase). We segmented each of the stimulation epochs from the overall data, alongside an equal-length segment of baseline ECoG immediately preceding the stimulation epochs. We analyzed a single representative 10 mA stimulation epoch from each subject, as these amplitudes created the strongest artifacts and thus represent a worst-case scenario for artifact suppression. We then used the segmented baseline and stimulation epochs to compare the performance of the PWNP and ICA artifact suppression algorithms.



**Figure 1.** MR-CT co-registered images with ECoG grids. Subject 3 and Subject 4 had  $4 \times 5$  grids implanted in the right temporal area and left frontal area, respectively. Highlighted electrodes mark the bipolar stimulation channels.

*2.1.3. MEA Data Collection* MEA data were collected from two  $7 \times 7$  sputtered iridium oxide film (SPIROF) tipped microelectrode arrays (Blackrock<sup>®</sup> Neurotech, Salt Lake City UT) implanted in the primary somatosensory cortex (S1) of a single human subject (Subject 5) with a C5-level complete spinal cord injury [12]. As no imaging data was available for this subject, the approximate location of this array on a brain template is shown in Figure 2. Data were collected at a sampling frequency of 30 kHz from a total of 96 channels, as one electrode on each array was designated as the reference/ground (Electrode 49 for MEA1 and Electrode 98 for MEA2). An ICMS device (Blackrock CereStim, Blackrock<sup>®</sup> Neurotech, Salt Lake City UT) delivered stimulation through a single electrode (Electrode 19). A one-second stimulation epoch consisted of delivering a train cathodic-leading biphasic square pulses at a frequency of 294 Hz, phase width of  $200 \mu\text{s}$ , and a current amplitude of  $100 \mu\text{A}$ . These stimulation epochs were repeated 10 times with an average of 27 seconds in-between. Electrodes that were saturated due to stimulation artifacts exceeding an absolute voltage amplitude of  $8192 \mu\text{V}$  (mostly occurring on MEA1, see Figure 2) were excluded from analysis as they contained no useful neural data, leaving 54 non-saturated channels across both arrays. Similarly to the ECoG data, we segmented stimulation epochs from the overall data. We then high-passed ( $\geq 0.1$  Hz) and linearly de-trended these stimulation epochs and subsequently appended them into one 10-second epoch. We also segmented a 10-second epoch of data occurring between two consecutive stimulation epochs and designated these data as the



**Figure 2.** Approximate locations of the two MEAs superimposed on an MRI brain template. Two 7 x 7 MEAs were implanted over S1. Stimulation was applied to electrode 19 on MEA1 (electrodes 1 - 49). Saturated electrodes are marked in red. Electrodes 49 and 98 were used as reference for their respective MEAs.

baseline epoch. These baseline and stimulation epochs were saved for later analysis.

## 2.2. Artifact Suppression Procedures

*2.2.1. Pre-whitening and Null Projection (PWNP) Algorithm* The PWNP algorithm is fully described in [38], and its theoretical basis can be found in [40]. Appendix A also gives a detailed, self-contained account of the main assumptions and mathematical derivations. The algorithm exploits the fact that artifacts are typically much stronger than neural activity, and thus reside in a low-dimensional subspace, corresponding to the highest singular values of the data matrix. This allows the artifact subspace to be readily identified, so the data can be projected to its orthogonal complement through null projection. The resulting data will then reside in an artifact-free subspace. The pre-whitening step serves to remove spatial correlations between electrodes, which improves the SNR of neural signals and the accuracy of the signal and artifact subspace estimates [41]. Mathematically (see Appendix A for derivations) these steps can be expressed as:

$$\mathbf{X}_S^{\text{clean}} = \Sigma_B^{\frac{1}{2}} \mathbf{H} \mathbf{H}^T \left[ \Sigma_B^{-\frac{1}{2}} (\mathbf{X}_S - \mu_S \mathbf{1}^T) \right] + \mu_S \mathbf{1}^T \quad (1)$$

where  $\mathbf{X}_S \in \mathbb{R}^{n \times t_S}$  represents the artifact-laden stimulation data from  $n$  channels over  $t_S$  time samples. Correspondingly,  $\mathbf{X}_S^{\text{clean}} \in \mathbb{R}^{n \times t_S}$  represents the stimulation data after it has been “cleaned” by the PWNP method. The vector  $\mu_S \in \mathbb{R}^{n \times 1}$  is the time average of  $\mathbf{X}_S$  and  $\mathbf{1} \in \mathbb{R}^{t_S \times 1}$  is a vector whose entries are 1, so that subtracting  $\mu_S \mathbf{1}^T$  de-means

$\mathbf{X}_S$ . The pre-whitening process is then completed by multiplying the de-measured data by the pre-whitening matrix  $\Sigma_B^{-\frac{1}{2}} \in \mathbb{R}^{n \times n}$ , which is estimated from the baseline data. The columns of the null-projection matrix,  $\mathbf{H} \in \mathbb{R}^{n \times (n-d)}$ , are the pre-whitened data's left singular vectors that correspond to its lowest  $n - d$  singular values, where  $d$  is the dimension of the artifact subspace. The artifact subspace dimension is equivalent to the number of singular values  $\sigma$  that satisfy the criterion  $\sigma > \alpha\sqrt{t_S - 1}$  ( $\alpha > 1$ ). This is modified from the original criterion  $\sigma \approx \sqrt{t_S - 1}$  to account for noise [38]. The reader is referred to Appendix A for a full derivation of this criterion. The pre-whitened data are then null-projected (pre-multiplied by  $\mathbf{H}^T$ ), reconstructed (pre-multiplied by  $\mathbf{H}$ ), and re-colored (pre-multiplied by  $\Sigma_B^{\frac{1}{2}}$ ). Finally the mean,  $\mu_S$ , is added back to the data.

To determine the optimal value of the threshold multiplier  $\alpha$  for each neural data set, we used the following procedure. First, the average power  $\bar{P}$  was calculated for each channel as:

$$\bar{P} = \frac{1}{n_f} \int_{f_1}^{f_2} PSD(f) df \quad (2)$$

where  $PSD(f)$  is the power spectral density,  $[f_1, f_2]$  is the relevant frequency range, and  $n_f$  is the number of frequency points in that range. For the narrow-band, 30 Hz, EEG stimulation, we used the frequency range of  $[f_1, f_2] = [29, 31]$  Hz. For the broad-spectrum ECoG/MEA stimulation response, we used a frequency range of  $[f_1, f_2] = [0, f_s/2]$ , where  $f_s$  is the sampling frequency. We then calculated  $\bar{P}$  for both stimulation ( $P_s$ ) and baseline ( $P_b$ ) data and identified a worst-case electrode as the electrode with the largest average power difference ( $P_s - P_b$ ). We subsequently varied  $\alpha = [1, \alpha_{max}]$ , where  $\alpha = 1$  is the theoretical optimum and  $\alpha_{max}$  is a conveniently chosen value at which the artifact subspace collapses into an empty set ( $d = 0$ ). For each value of  $\alpha$  within this range, we removed the artifacts from the worst-case electrode according to Equation 1, and calculated its average power  $P_\alpha^{\text{clean}}$ . We then find the optimal value  $\hat{\alpha}$  by:

$$\hat{\alpha} = \arg \min_{\alpha} |P_\alpha^{\text{clean}} - P_b|, \quad \alpha \in [1, \alpha_{max}] \quad (3)$$

Note that this procedure takes advantage of the fact that the power difference between the baseline and stimulation data is largely due to the presence of artifacts. Therefore, upon cleaning, we expect that  $P_{\hat{\alpha}}^{\text{clean}}$  approaches  $P_b$ .

*2.2.2. ICA Artifact Suppression Procedure* To benchmark the artifact suppression performance of the PWNP method, we compared it against an ICA-based back-projection method. We used the FastICA toolbox in MATLAB [42] to generate independent components (ICs) for the EEG, ECoG, and MEA data. Unlike the PWNP method, in which the artifact subspace is readily identifiable due to rank-sorting of components, the ICA method necessitates a combinatorial search to identify the optimal artifact subspace. Given the number of channels we had in EEG, ECoG and MEA, it was not practical to perform a full combinatorial search. Instead, we used the

following heuristic approach. Converged ICs were inspected in the time and frequency domains to identify components containing artifact features (e.g. patterns occurring at the pulse train frequency, power peaks occurring at the pulse train frequency and super-harmonics [16, 18]). After the components comprising the artifact subspace were identified and nulled, the cleaned data were obtained via the back-projection method [43]. This procedure was repeated by nulling various combinations of artifact components until we achieved the best suppression result, defined as the minimum average power difference between ICA-cleaned and baseline states on the worst-case electrode.

### 2.3. Artifact Suppression Evaluation Analyses

*2.3.1. EEG Artifact Suppression Evaluation* Using the procedures for PWNP (Section 2.2.1) and ICA (Section 2.2.2), we cleaned the stimulation EEG data of Subjects 1 and 2. Subsequently, we calculated the change in SNR and SIR due to artifact suppression for both PWNP and ICA methods in order to compare their performances. Due to the eyes open/closed task, we expect to observe modulation in the  $\alpha$  band (8 – 12 Hz) [44]. To quantify this modulation, we introduced a SNR-like metric, which compares the separability of the eyes-open/eyes-closed states. Specifically, we calculated SNR as a deflection coefficient [45]:

$$\text{SNR}(f) = 10 \log \sqrt{\frac{(\mu_c(f) - \mu_o(f))^2}{0.5(\sigma_c^2(f) + \sigma_o^2(f))}}, f \in [8, 12] \text{ Hz} \quad (4)$$

where  $\mu_c(f)$  and  $\mu_o(f)$  represent the average power in the  $\alpha$  band over eyes-closed and eyes-open epochs, respectively. Correspondingly,  $\sigma_c^2(f)$  and  $\sigma_o^2(f)$  are the eyes-closed and eyes-open power variances in the  $\alpha$  band. The overall SNR was then calculated as the average of  $\text{SNR}(f)$  over the  $\alpha$  band.

To further compare suppression performances, we also introduced a metric to represent the SIR. This metric allowed us to evaluate the degree of stimulation interference in comparison to the amount of occipital  $\alpha$ -band modulation. We calculated the SIR as the ratio of the maximum average power in the  $\alpha$  band to the maximum average power in the interference band, which we chose to be 29 - 31 Hz to capture the narrow-band, 30 Hz, sine stimulation (Equation 5). We evaluated SIR during eyes-closed segments to ensure that the occipital  $\alpha$  waves were present, i.e.:

$$\text{SIR} = 10 \log \frac{\max_{8 \leq f \leq 12} \mu_c(f)}{\max_{29 \leq f \leq 31} \mu_c(f)} \quad (5)$$

For each channel, we calculated the SNR and SIR before suppression and after application of either the PWNP or ICA method. We then characterized the suppression performance of each method by calculating  $\Delta\text{SNR}$  and  $\Delta\text{SIR}$ , respectively defined as the change in SNR and SIR due to artifact suppression. We expect that successful

artifact suppression would increase the SIR ( $\Delta\text{SIR} > 0$ ) while not decreasing SNR ( $\Delta\text{SNR} \geq 0$ ). We subsequently used a signed rank test to confirm whether SIR values across electrodes were significantly different before and after applying either suppression method. The same test was also used to assess whether the  $\Delta\text{SIR}$  values across electrodes were significantly different between the two suppression methods. Finally, to visualize and compare the suppression results spatially, the  $\Delta\text{SIR}$  values were color-coded and overlaid on a topographical EEG map.

*2.3.2. ECoG Artifact Suppression Evaluation* ECoG stimulation data were cleaned using the PWNP and ICA methods described in Section 2.2.1 and 2.2.2, respectively. Unlike those in the EEG protocol, the subjects in the ECoG protocol did not perform any behavioral tasks, therefore an SNR could not be defined as was done with the EEG data. Instead, we evaluated the artifact suppression performance by comparing the PSD of cleaned data to that of the baseline. This approach is consistent with our previous assertion that most of the power differences between stimulation and baseline data are due to artifacts (see Section 2.2.1). For example, significant spectral differences between cleaned and baseline data, especially at the stimulation frequency and its super-harmonics, would indicate inadequate suppression. Conversely, more aggressive suppression could potentially remove neural features in addition to artifacts, resulting in “over-cleaning”. To control for this outcome, we performed a baseline control experiment where we applied both artifact suppression methods to the baseline epoch. Since baseline data contain no artifacts, we expect the baseline epoch to remain relatively unaffected by this procedure. To visualize potential distortions due to either method of artifact suppression, we compared the baseline epoch before and after cleaning in the time and frequency domains. We quantified the distortion in the time domain by calculating the root-mean-squared error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (b_i - c_i)^2} \quad (6)$$

where  $n$  is the number of samples in the baseline epoch time series,  $b$  is original baseline data and  $c$  is the same data after suppression. These RMSE values were then color-coded and mapped to images of the ECoG grids to spatially visualize the effect of cleaning over multiple electrodes. Additionally, time domain data for a representative channel were shown before and after the PWNP/ICA cleaning procedures. We then visualized the effects of the cleaning procedures in the frequency domain by plotting average PSDs of baseline and cleaned epochs. A signed rank test was then performed to determine whether the power distributions were significantly different across frequencies. We note that the PWNP method may have an unfair advantage in this comparison, as it was trained using the baseline epoch. To control for this, we tested the performance of both artifact suppression methods on an additional 100 baseline epochs, segmented from ECoG data occurring outside of stimulation periods. These tests and their results are detailed in Appendix B.

We then examined the neural time series for the stimulation, PWNP-cleaned, ICA-cleaned and baseline conditions in order to assess the signal quality after artifact suppression. We next examined the aforementioned conditions in the frequency domain by calculating the PSD. We accomplished this by splitting the epoch for each condition (stimulation, PWNP-cleaned, ICA-cleaned, baseline) into 10 equal-length subsections, and then performing the fast Fourier transform on each section to obtain their PSDs. The average and standard deviation of these PSDs were then calculated over the sections. We subsequently plotted the PSDs for each condition for the worst-case electrode, defined as in Section 2.2.1. This allowed us to observe the frequency domain features introduced to the baseline PSD by the stimulation. Additionally, these PSDs allowed us to assess the reduction of artifact features by both suppression methods.

Unlike EEG artifacts, which had a narrow-band frequency response, ECoG artifacts had a broadband power distribution. As such, the SIR defined by Equation 5 could not be used, so we introduced a separate SIR-like metric to quantify the artifact suppression performance. As was described in Section 2.2.1, we expect the power distribution of cleaned data to approach that of the baseline. To quantify the separation between power distributions, we calculated a variant of the deflection coefficient [46], which we refer to as the interference index. Note that this is a general metric that is capable of accounting for overlapping means and unequal variances [47]. For two power distributions  $a$  and  $b$ , the interference index at frequency  $f$  is defined as:

$$I(f) = \frac{1}{2} \log \frac{\sigma_t^2(f)}{\sigma_a(f)\sigma_b(f)} \quad (7)$$

where  $\sigma_a(f)$  and  $\sigma_b(f)$  are the standard deviations of PSDs for two conditions. Since the PSD sample sizes for each condition are equal ( $n = 10$ , see previous paragraph), the total standard deviation,  $\sigma_t(f)$ , can be expressed as:

$$\sigma_t^2(f) = \frac{\sigma_a^2(f) + \sigma_b^2(f)}{2} + \frac{[\mu_a(f) - \mu_t(f)]^2}{2} + \frac{[\mu_b(f) - \mu_t(f)]^2}{2} \quad (8)$$

where  $\mu_a(f)$  and  $\mu_b(f)$  are the means of the two PSDs and  $\mu_t(f) = \frac{1}{2}(\mu_a(f) + \mu_b(f))$  is the total mean. We first calculated the interference index between stimulation and baseline conditions to serve as a positive control (a = stimulation, b = baseline). Subsequently, we calculated the interference index between the cleaned and baseline conditions (a = cleaned, b = baseline) to evaluate the effectiveness of each artifact suppression method. Note that smaller artifact interference will result in a lower value of  $I(f)$  (inverse to SIR). Ideally, in the case where there are no artifacts, the interference index will approach zero ( $\sigma_a(f) = \sigma_b(f)$  and  $\mu_a(f) = \mu_b(f)$ ). We then plotted the interference indices for the stimulation, PWNP-cleaned, and ICA-cleaned conditions for the worst-case electrode.

We further qualified each interference index by performing a rank-sum test to identify frequencies with power distributions different than those of the baseline condition. These significant interference frequencies were then marked on the

interference index plots. We expect the superior suppression method to result in fewer significant interference frequencies. As a positive control, we also calculated the number of significant interference frequencies in the stimulation data.

We additionally characterized the suppression results by calculating the interference indices for all electrodes in the grid between different conditions for each ECoG electrode. Specifically, we first calculated the interference index between the stimulation and baseline conditions as a positive control. Subsequently, we calculated the interference indices between the cleaned and baseline conditions to quantify the residual artifact for both artifact suppression methods. We used signed rank test to establish the statistical significance of these results across all electrodes. Subsequently, we used the same test to compare the performances of PWNP and ICA interference indices. To visualize the distribution of residual artifacts, we spatially interpolated, color-coded, and mapped the interference indices onto MR-CT co-registered images. For both subjects, we used pre-implantation MR images and post-implantation CT images to co-register ECoG grids onto a 3D brain rendering. We followed the same co-registration process outlined in [18].

Finally, since clinically obtained stimulation data by definition has no ground truth, we sought to simulate artifact data wherein the underlying neural signals are known. Therefore, we extracted artifacts from the stimulation epoch, and overlaid the average artifact waveforms onto the 100 baseline epochs (as explained above). We then applied the artifact suppression methods to these simulated data, and compared the results to the ground truth signals to quantify the performance in the time and frequency domains (see Appendix B).

*2.3.3. MEA Artifact Suppression Evaluation* MEA stimulation data were cleaned using the PWNP and ICA methods described in Section 2.2.1 and 2.2.2, respectively. Similar to the ECoG protocol, the subject in the MEA protocol did not perform any behavioral tasks, therefore we evaluated the artifact suppression performance by comparing the PSD of cleaned data to that of the baseline. To rule out “over-cleaning”, we again performed baseline control experiments and compared the baseline epoch before and after cleaning in the time domain (RMSE) and frequency domain (PSD). Similar to the ECoG data, we tested the performance of both artifact suppression methods on an additional 100 baseline epochs, segmented from MEA data occurring outside of stimulation periods. These tests and their results are detailed in Appendix C.

We then examined the MEA time series for the stimulation, PWNP-cleaned, ICA-cleaned and baseline conditions in order to assess the signal quality after artifact suppression. Similarly to the ECoG data, these conditions were analyzed in the frequency domain for the worst-case electrode. Specifically, we calculated the PSDs and interference indices across these conditions to account for the broadband nature of the MEA stimulation artifacts. We also used the rank-sum test to identify those frequencies exhibiting significantly different power distributions between the baseline and cleaned data.

To characterize the artifact suppression results beyond the worst-case electrode, we

calculated the frequency-averaged interference index for each non-saturated electrode of the MEA. These interference indices were then spatially interpolated, color-coded and mapped onto the coordinates of the MEA superimposed onto a 3D brain render. Note that for Subject 5, brain images were not available. Instead, the location of the MEA was estimated based on photographs of the implantation location taken during surgery and aligned to a template brain using anatomical landmarks. To compare the significance of these results across all electrodes and between PWNP and ICA, we used a signed rank test to establish the statistical significance between conditions across all electrodes.

Finally, in the absence of a behavioral task in the MEA protocol, we wanted to evaluate the ability of both artifact suppression methods to retain action potentials in the MEA data. Specifically, we identified an electrode exhibiting action potentials during the baseline condition. We then counted the number of action potentials occurring during the stimulation epochs for that electrode. After artifact suppression, we calculated the fraction of retained action potentials for both the PWNP-cleaned and ICA-cleaned data.

### 3. Results

#### 3.1. EEG Artifact Suppression Results

Artifact suppression results for EEG data showed that both methods successfully attenuated the artifacts, as evidenced by SIR improvements, while preserving the SNR. Using the PWNP method, we estimated the dimension of the artifact subspace to be four ( $d = 4$ ,  $\hat{\alpha} = 2.9$ ) for Subject 1 and  $d = 7$  ( $\hat{\alpha} = 1.2$ ) for Subject 2. For the ICA method, 3 of 19 converged components were identified as artifact components for Subject 1, and 3 of 15 were identified for Subject 2. The components comprising the artifact subspace in PWNP and the artifact components in ICA all contained significant power at the stimulation frequency (30 Hz). Figure 3 shows the change in the SNR (Equation 4) and SIR (Equation 5) after applying each artifact suppression technique. As expected, both methods resulted in SIR improvements that were statistically significant (a paired, right-tailed, signed rank test; Subject 1,  $p = 0.00007$  for both methods; Subject 2,  $p = 0.00007$  for both methods). Comparing between methods, the  $\Delta$ SIR was greater for the PWNP method than for the ICA method for both subjects. For example, for Subject 1, the best channel (C3) exhibited a  $\Delta$ SIR of 47.2 dB for PWNP compared to 41.1 dB for ICA (see Table 1). This table also shows the summary statistics for  $\Delta$ SIR and  $\Delta$ SNR across all 19 channels for both subjects. Comparing the median  $\Delta$ SIR values between suppression methods, PWNP outperformed ICA by 7.04 dB in Subject 1 and 10.84 dB in Subject 2. Furthermore, these performance improvements were statistically significant for both subjects (paired, right-tailed, signed rank test; Subject 1,  $p = 0.00201$ ; Subject 2,  $p = 0.0016$ ). The spatial distribution of  $\Delta$ SIR (Figure 4) further highlights these results. Additionally, PWNP appeared to be the most effective on those electrodes

**Table 1.** Change in SIR and SNR after suppressing artifacts in EEG data for Subjects 1 and 2. The maximum  $\Delta$ SIR and  $\Delta$ SNR correspond to the electrodes that exhibited the largest SIR and SNR change upon artifact suppression. The median  $\Delta$ SIR and  $\Delta$ SNR are also reported with the median absolute deviation (MAD).

(dB)	Subject 1		Subject 2	
	PWNP	ICA	PWNP	ICA
max( $\Delta$ SIR)	47.20	41.10	43.08	35.20
max( $\Delta$ SNR)	4.41	3.51	5.13	7.75
median( $\Delta$ SIR)	34.22 $\pm$ 9.02	27.18 $\pm$ 8.34	31.79 $\pm$ 7.85	20.95 $\pm$ 7.04
median( $\Delta$ SNR)	0.18 $\pm$ 1.02	0.59 $\pm$ 0.84	-0.03 $\pm$ 1.49	0.34 $\pm$ 1.91

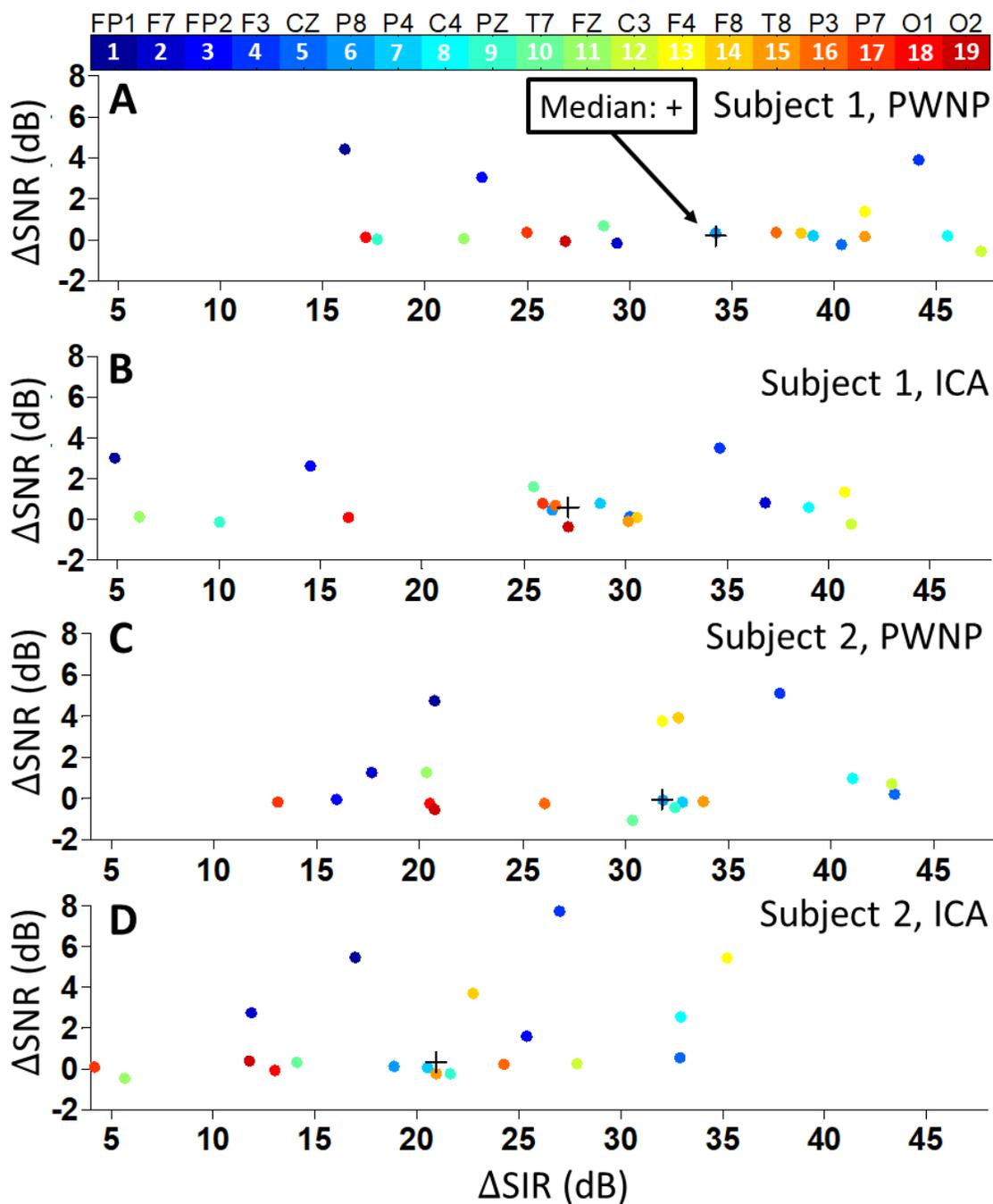
closest to the stimulation channel, which were the most severely affected by artifacts. In comparison to the SIR changes, SNR changes were generally small ( $<1$  dB). Statistical analysis showed that the SNR improvement after PWNP was statistically significant for Subject 1 (paired, right-tailed, signed rank test;  $p = 0.00258$ ), but not for Subject 2 ( $p = 0.14312$ ). For ICA, both subjects had statistically significant improvement in SNR (Subject 1,  $p = 0.00530$ ; Subject 2,  $p = 0.00136$ ). Note that SNR improvements are not the primary objective of artifact suppression.

### 3.2. ECoG Artifact Suppression Results

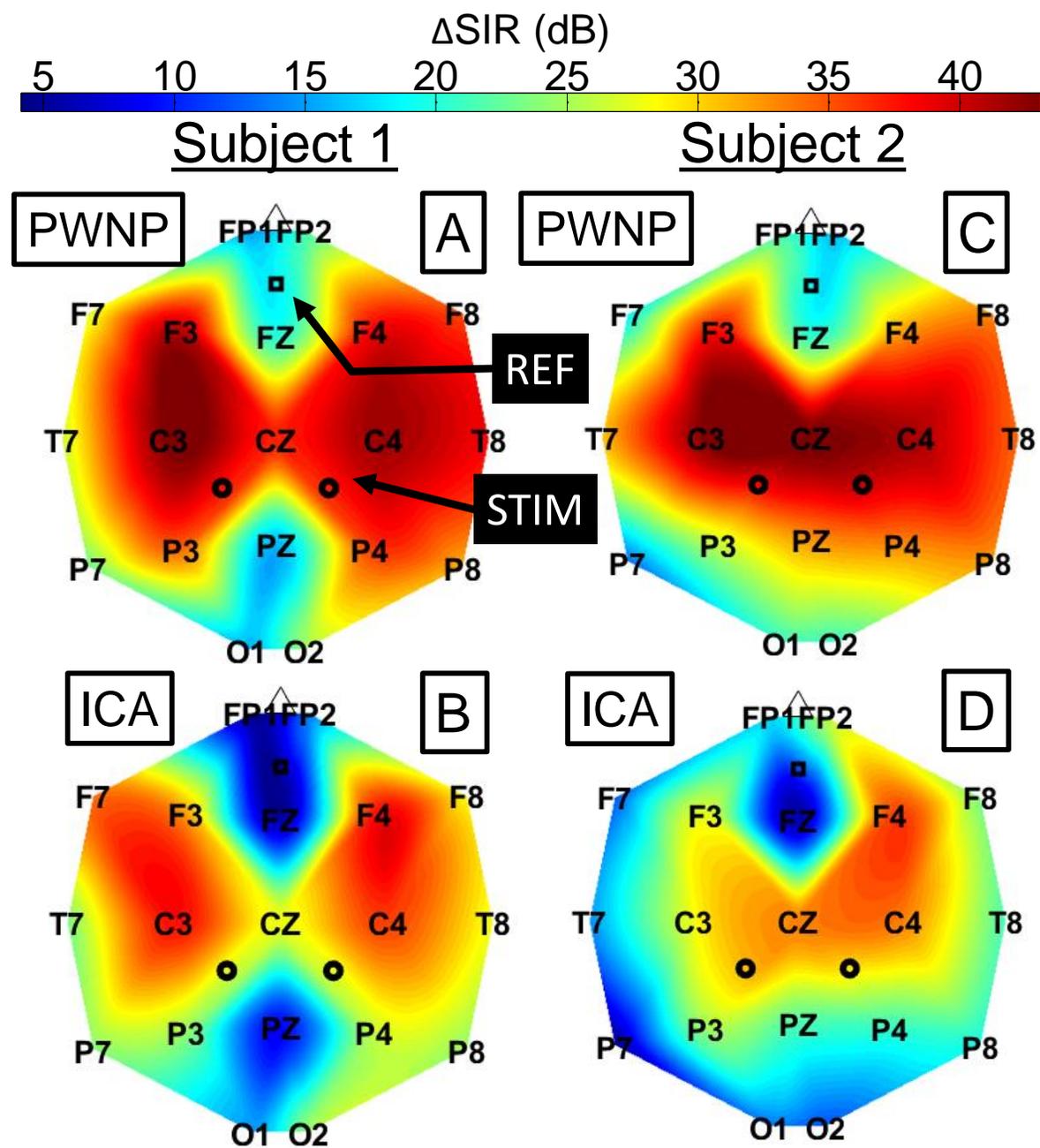
Like with EEG data, both methods were able to remove artifacts in the ECoG stimulation data. For the PWNP algorithm, we estimated the artifact dimension to be  $d = 12$  ( $\hat{\alpha} = 1.1$ ) for Subject 3 and  $d = 11$  ( $\hat{\alpha} = 1.1$ ) for Subject 4. For the ICA procedure, 10 of 18 components were identified as artifact components for Subject 3, while 11 of 15 were identified for Subject 4. The components comprising the artifact subspace in PWNP and the artifact components in ICA were nulled as explained in Sections 2.2.1 and 2.2.2, respectively.

ECoG signals exhibited prominent artifacts during stimulation. Figure 5A shows a representative segment of a stimulation epoch to illustrate this phenomenon. Despite their broadband nature, these artifacts were substantially reduced by both PWNP and ICA methods, examples of which are shown in Figure 5B and C, respectively. Generally, the signal amplitudes upon cleaning were more similar to those of baseline data (Figure 5D).

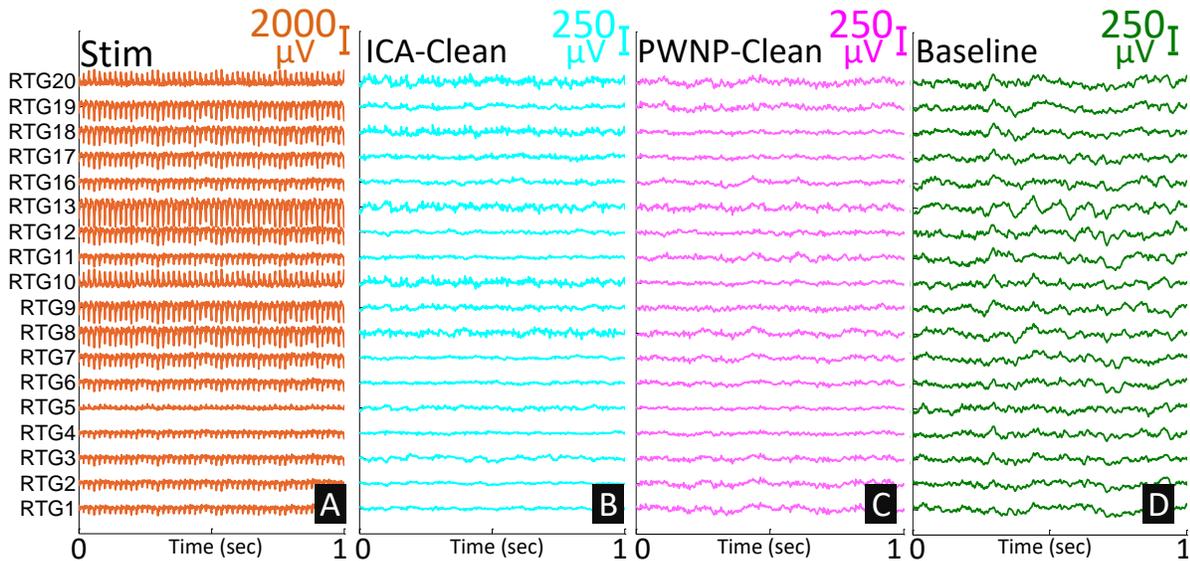
To illustrate the effectiveness of artifact suppression in the frequency domain, Figure 6 shows example PSDs before and after cleaning from the worst-case electrodes. Unsurprisingly, these were the electrodes closest to the stimulation channel that were co-linear with the moment of the stimulation dipole [16, 18]. In comparison to the baseline data, stimulation PSDs exhibited peaks at the stimulation frequency (50 Hz) and its super-harmonics, as well as a broadband increase. Upon artifact suppression, these artifact features were largely reduced and the PSDs were brought closer to the



**Figure 3.** Plots of the change in SIR and SNR values upon applying PWNP/ICA artifact suppression techniques for the EEG data for Subjects 1 and 2. Each color corresponds to a different EEG channel. The median  $(\Delta\text{SIR}, \Delta\text{SNR})$  point across electrodes is indicated on each plot by a “+”. The SIR improved significantly for both subjects, with the PWNP method outperforming the ICA method. Both methods preserved the SNR, as evidenced by median  $\Delta\text{SNR} \approx 0$ . A numerical summary of these results is provided in Table 1.



**Figure 4.** Maps of  $\Delta\text{SIR}$  for both subjects for the PWNP-cleaned (A,C) and ICA-cleaned (B,D) data. Black square (reference electrode), black circles (stimulation electrodes). PWNP outperforms ICA for both subjects, with the most substantial suppression occurring on the most contaminated channels (C3, Cz, C4). Hot/cold spots away from electrodes are spatial interpolation artifacts due to sparse electrode coverage.



**Figure 5.** Representative ECoG time series from 18 channels of the right temporal grid (RTG) of Subject 3 (see Figure 1). The stimulation channel (RTG14-15) is omitted due to saturation. (A) One-second segment from the stimulation epoch. (B) Same segment after PWNP artifact suppression. (C) Same segment after ICA artifact suppression. (D) One-second segment from a baseline epoch occurring immediately before the stimulation epoch. Note that the stimulation data is shown at an eighth of the scale of the others.

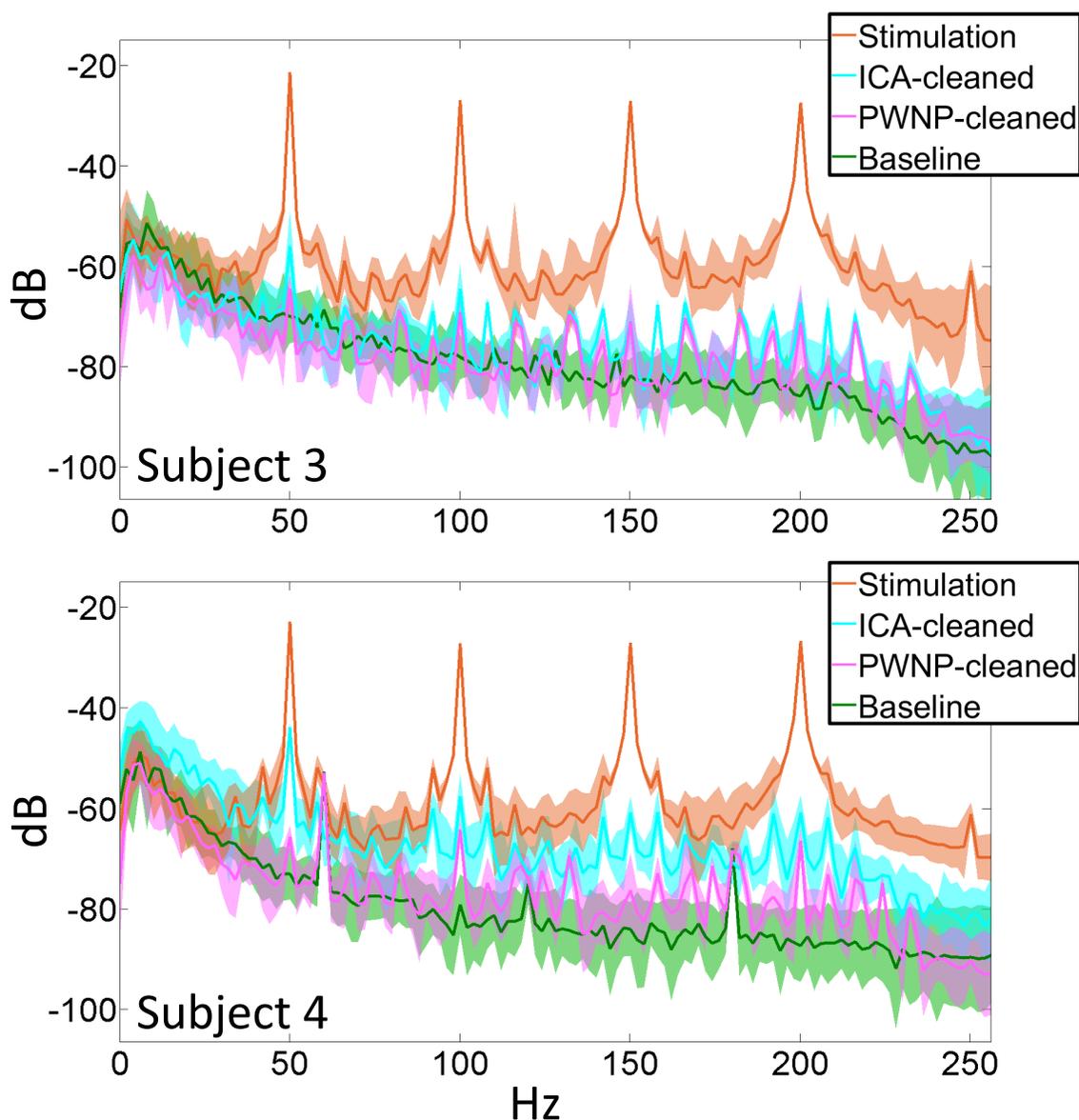
**Table 2.** Frequency-averaged interference indices,  $\bar{I}$ , (see Equation 7) and corresponding standard deviation,  $\sigma_I$ , for stimulation, PWNP-cleaned and ICA-cleaned conditions for worst-case electrode in ECoG data.

	Subject 3			Subject 4		
	Stim.	PWNP	ICA	Stim.	PWNP	ICA
$\bar{I}$	0.399	0.089	0.108	0.403	0.080	0.199
$\sigma_I$	0.318	0.102	0.140	0.320	0.100	0.121

baseline. Furthermore, the PSDs of PWNP-cleaned data were generally closer to the baseline in comparison to their ICA counterparts.

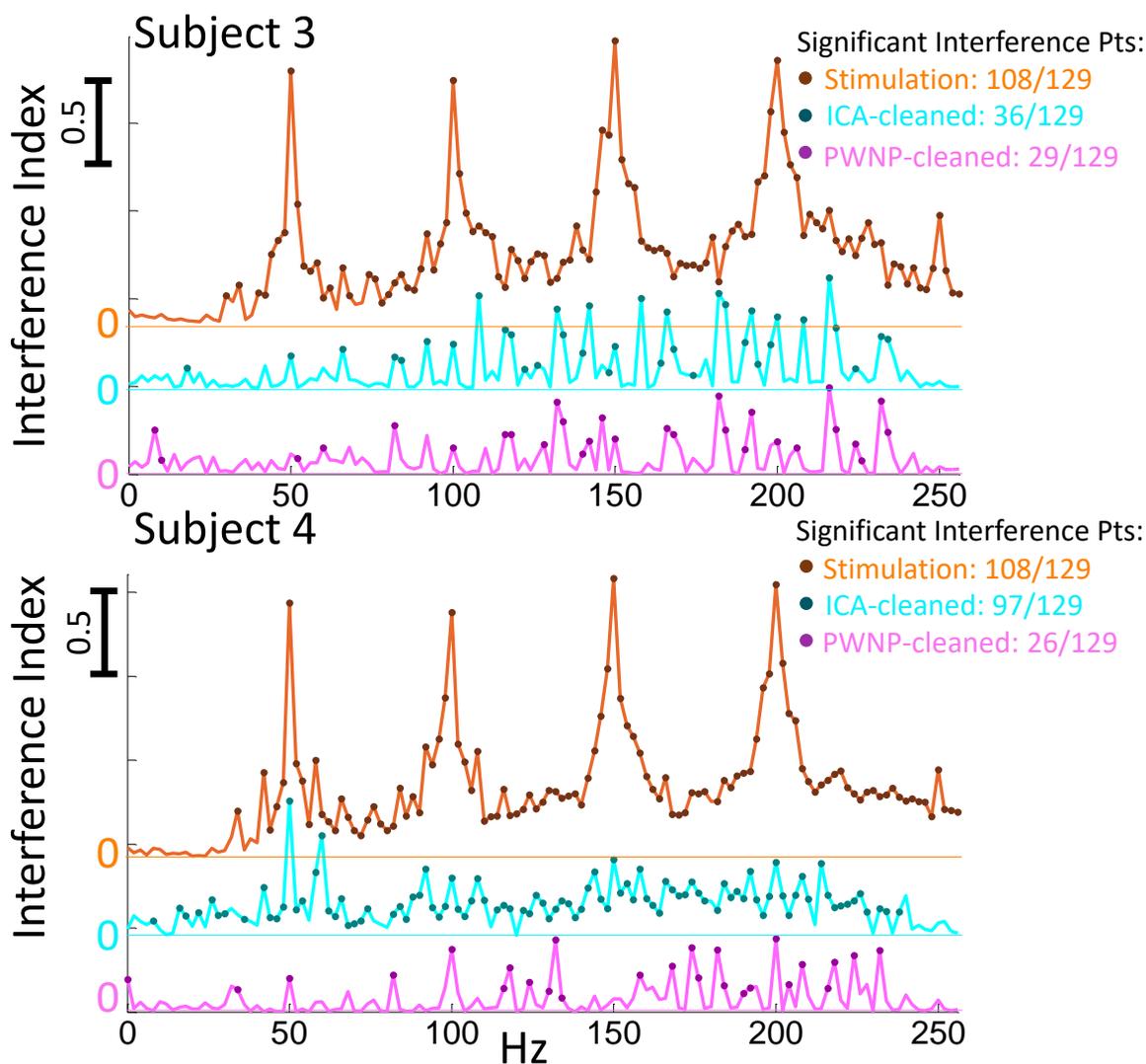
The interference index of the stimulation data exhibited peak values at the stimulation frequency (50 Hz) and its super-harmonics, similarly to stimulation data PSDs. Figure 7 illustrates this effect for the worst-case electrode. These peaks were largely removed after artifact suppression, with the PWNP method outperforming ICA. This was evidenced by PWNP achieving lower interference index values compared to ICA. Summary statistics for these results are given in Table 2. Moreover, PWNP-cleaned data had fewer frequencies with significant residual interference compared to ICA-cleaned data (rank-sum test  $p < 0.01$ ).

To demonstrate the effectiveness of artifact suppression methods beyond the worst-case electrode, we spatially mapped the interference indices (Figure 8). As expected,



**Figure 6.** PSDs of ECoG signals under four different conditions from a worst-case electrode from both subjects (RTG13 for Subject 3 and LFG13 for Subject 4, cf. Figure 1). Solid lines represent the PSD averages taken over 10 subsections and shades represent corresponding one standard deviation bounds.

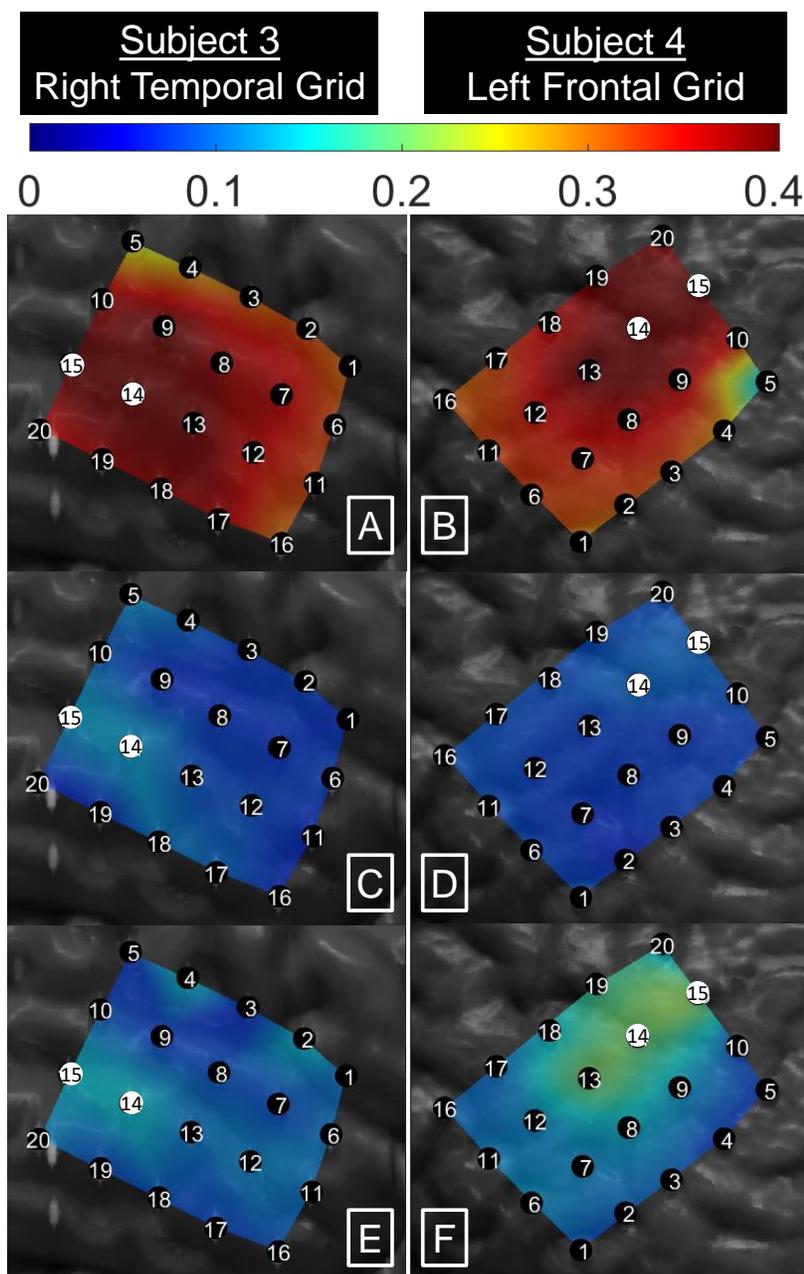
the map corresponding to the stimulation data exhibited the highest values, especially in the vicinity of the stimulation channel. Both artifact suppression methods effectively reduced the interference indices across electrodes. Specifically, the interference indices upon PWNP suppression became significantly smaller (paired, left-tailed, signed rank test: Subject 3,  $p = 4.8 \times 10^{-5}$ ; Subject 4,  $p = 1.1 \times 10^{-4}$ ). Similar behavior was observed after ICA suppression as well (Subject 3,  $p = 4.8 \times 10^{-5}$ ; Subject 4,  $p = 1.1 \times 10^{-4}$ ). Consistent with our worst-case electrode analyses, we observed that PWNP generally outperformed ICA, particularly on the electrodes closest to the stimulation channel.



**Figure 7.** Interference indices calculated for worst-case electrodes from both subjects. Filled-circle markers indicate frequencies with significant interference, as determined by rank-sum test ( $p < 0.01$ ). The PWNP method achieved superior suppression results, as it generally resulted in lower interference indices, as well as fewer frequencies with significant residual interference.

Specifically, the PWNP method yielded lower interference indices across electrodes (paired, left-tailed signed rank test; Subject 3:  $p = 0.010495$ ; Subject 4:  $p = 0.000994$ ).

Figure 9 shows the results for the baseline control experiment. Since baseline data contained no artifacts, we expect artifact suppression methods to yield small RMSE values between baseline and artifact-suppressed baseline data. For PWNP, the average RMSE value across electrodes was  $15.3 \pm 2.3 \mu\text{V}$ , which accounted for only 6% of the pre-cleaning baseline voltage swing ( $256 \mu\text{V}$ ). In contrast, approximately four times higher RMSE values were obtained with the ICA method ( $68.8 \pm 27.4 \mu\text{V}$ ), suggesting that it imposed more significant signal distortions. To visualize these distortions,



**Figure 8.** Frequency-averaged interference indices (Equation 7) were spatially interpolated, color-coded and mapped to cortical surfaces. For the saturated stimulation electrodes (white), the values were imputed to the highest value in the map to preserve the continuity of the interpolation. (A,B) Interference indices for stimulation data. (C,D) Interference indices for PWNP-cleaned data. (E,F) Interference indices for ICA-cleaned data. Note that the stimulation electrodes were saturated, so their value was imputed to the highest in the grid to preserve interpolation continuity.

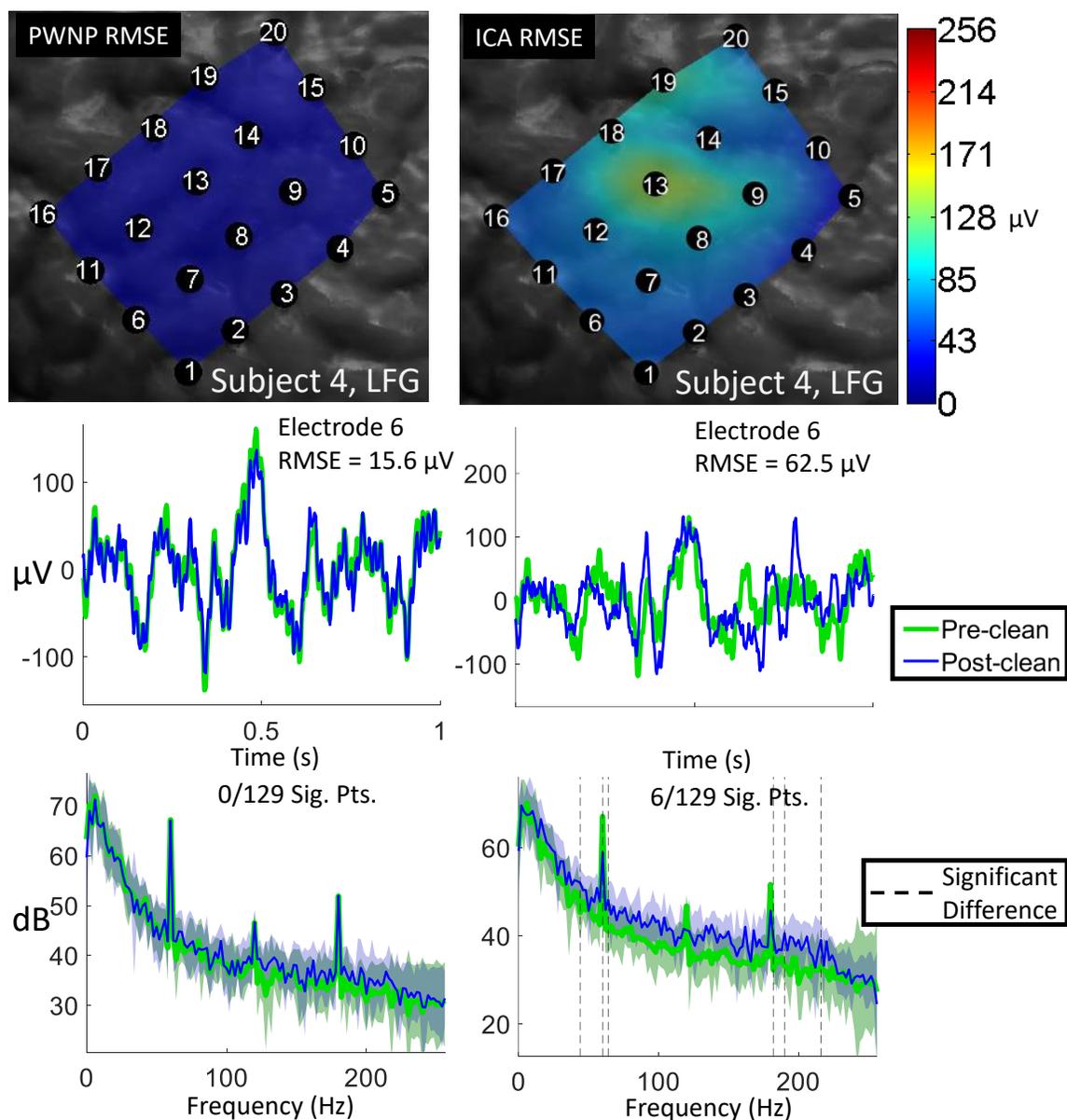
Figure 9 also shows time domain baseline signals for a representative electrode before and after artifact suppression. We selected the representative electrode as the electrode exhibiting the RMSE closest to the median RMSE across the grid. As evidenced by these examples, the PWNP method introduced much less distortion in the time domain compared to the ICA method. We additionally characterized these post-suppression baseline distortions in the frequency domain. Comparing the PSDs before and after artifact suppression (signed rank test,  $p < 0.01$ ), we identified no frequencies exhibiting significantly different power distributions after PWNP artifact suppression, while 6 (out of 129) frequencies were identified for ICA. The PWNP method may have benefited over the ICA method from the fact that the baseline data epoch in this control experiment had been used to calculate the pre-whitening matrix,  $\Sigma_B^{-\frac{1}{2}}$ , (see Equation 1). To rule this out, we performed control experiments on additional baseline epochs (Appendix B), while retaining the same PWNP and ICA parameters. Specifically, we found that PWNP yielded an average RMSE that accounted for only 5.5% of the baseline voltage swings. The distortions due to ICA were three times as high, with the average RMSE value reaching 16.5% of the baseline voltage swing. In the frequency domain, we found that PWNP on average yielded 1.9% of frequencies with significantly different power distribution compared to 43.6% of the frequencies yielded by ICA. For a detailed account of these results and supporting figures, the reader is referred to Appendix B.

The advantages of the PWNP artifact suppression method over its ICA counterpart were retained with the simulated artifact data. Specifically, upon PWNP artifact suppression, the residual signals generated an average RMSE accounting for 5.6% of the baseline voltage range. Similarly to the baseline control experiments, the average distortions due to ICA were three times as high (16.7%). Furthermore, PWNP on average yielded 4.5% of frequencies with power distribution significantly different from those of the corresponding ground truth signals. For ICA, this fraction was 46.5%, which is also in line with the baseline control experiments. Appendix B provides a detailed account of these results.

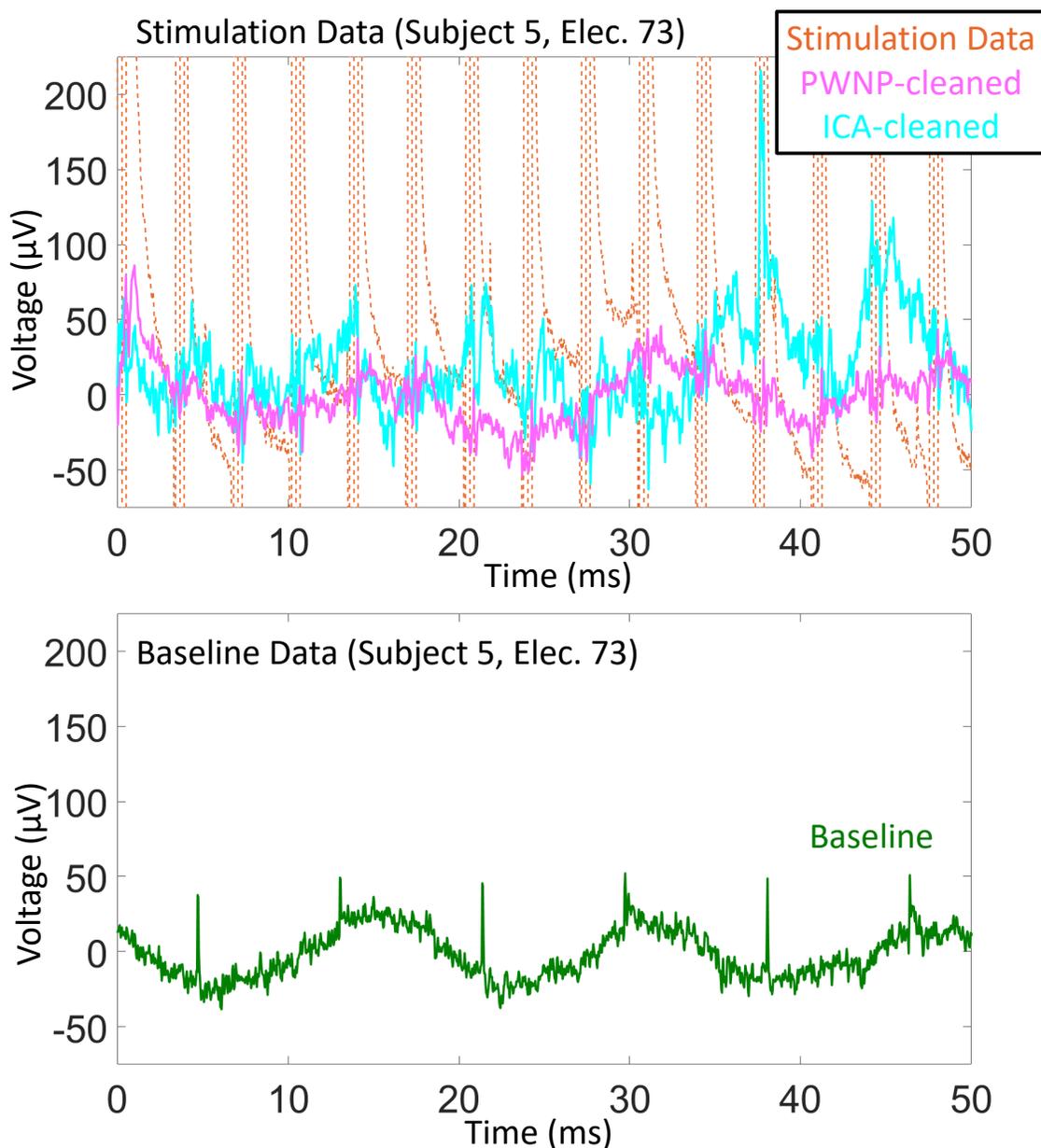
### *3.3. MEA Data Artifact Suppression Results*

Artifact suppression evaluation for the MEA data yielded similar results as for ECoG data, with the PWNP method generally demonstrating superior suppression results. After excluding saturated electrodes (artifact amplitudes exceeding 8.7 mV) from both MEAs, we used data from the remaining 54 channels (cf. Figure 2) to train the PWNP and ICA algorithms. For the PWNP method, we estimated the artifact subspace dimension to be  $d = 33$  ( $\hat{\alpha} = 1.2$ ). For the ICA method, we identified 49 ICs as artifacts (out of 54 converged components). The artifact subspace in PWNP and the artifact components in ICA were nulled as explained in Sections 2.2.1 and 2.2.2, respectively. Since most of the electrodes on MEA1 were saturated, we focus on results from MEA2.

Figure 10 shows a representative segment of the stimulation data from the worst-case electrode on MEA2. Evidently, neural signals were dominated by extremely strong

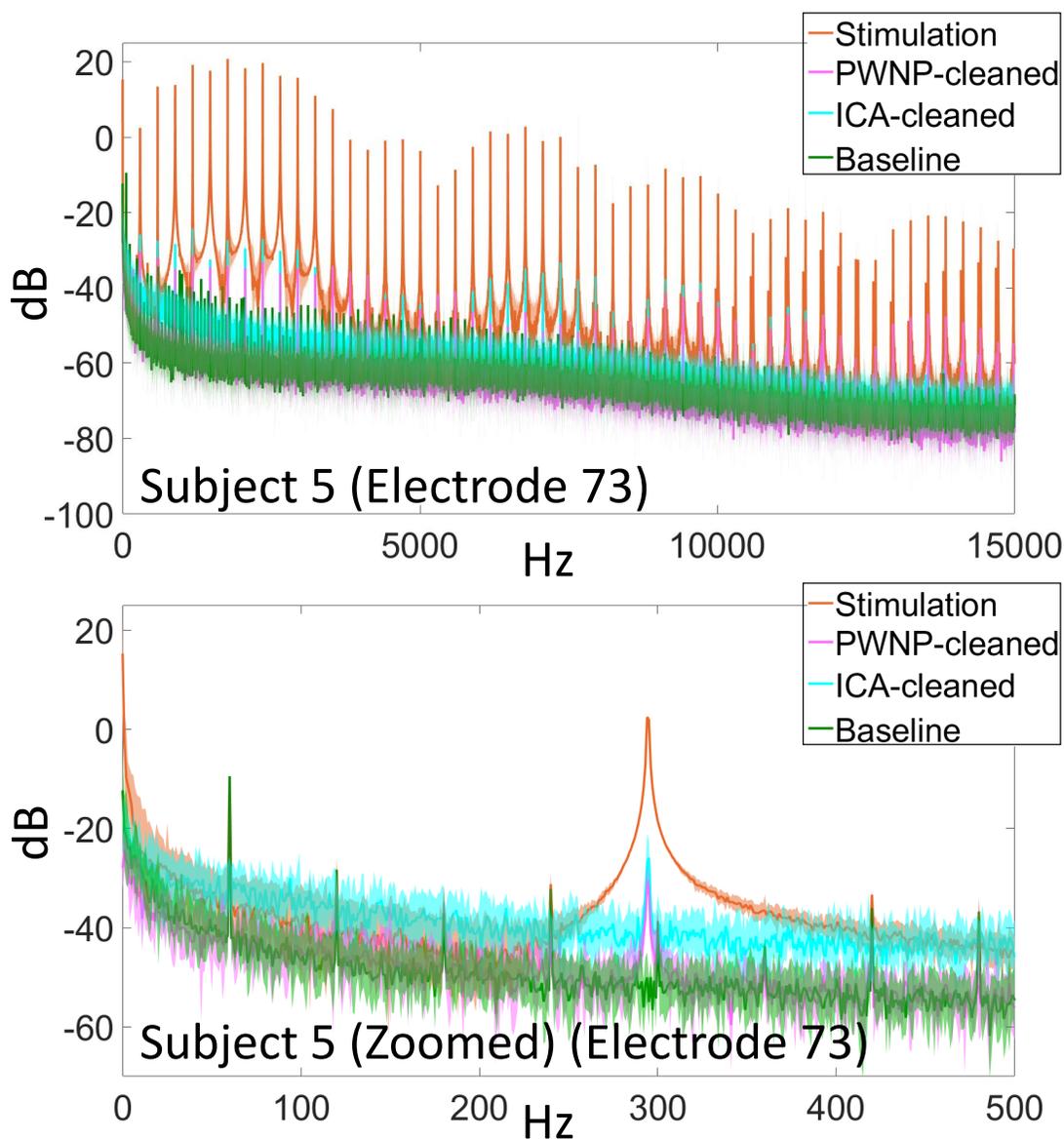


**Figure 9.** Baseline control experiment results for PWNP (Left) and ICA (Right) methods. (Top) RMSE values (Equation 6) spatially interpolated, color-coded, and mapped to MR-CT co-registered images of the ECoG grids from Subject 4. Color bar range is from  $0 \mu\text{V}$  to  $256 \mu\text{V}$  (maximum absolute voltage of the baseline data). (Middle) Representative baseline time domain examples before and after artifact suppression (Bottom) Representative frequency domain examples (mean PSD) from the same electrode before/after artifact suppression. Dashed lines indicate frequencies where the power distribution significantly differed before and after artifact suppression (signed rank test,  $p < 0.01$ ).



**Figure 10.** (Top) A representative 50-ms segment of the stimulation epoch from the worst-case electrode on MEA2, along with the same data after artifact suppression. To preserve the scale, the stimulation data have been de-meant and truncated. (Bottom) A duration-matched baseline segment immediately preceding the stimulation data.

artifacts, whose amplitudes exceeded those of baseline signals by as much as two orders of magnitude. This is in contrast to the ECoG artifacts, which were generally an order of magnitude larger than the corresponding baseline signals (see Figure 5). Despite their large amplitude and broadband power distribution, these artifacts were still substantially reduced by both PWNP and ICA methods, bringing the signal amplitudes closer to those of the baseline data.



**Figure 11.** PSDs of signals under four different conditions from the worst-case electrode (Electrode 73) of MEA2. Solid lines represent the PSD averages taken over 10 stimulation epochs and shades represent corresponding one standard deviation bounds. (Top) Full PSD. (Bottom) Same PSD, zoomed to local field potential frequency range.

Figure 11 shows these representative signals in the frequency domain. The stimulation data PSDs exhibited a broadband increase, as well as peaks at the stimulation pulse train frequency (294 Hz) and its super-harmonics (589 Hz, 883 Hz, 1178 Hz, etc.). Note that these frequencies were rounded to the nearest whole number due to the 1-Hz frequency resolution of the PSD. Both PWNP and ICA reduced the artifact-related spectral features in the stimulation data and brought the resulting PSDs closer to those of the baseline data. However, PWNP appeared to outperform ICA as it produced data with less residual artifact.

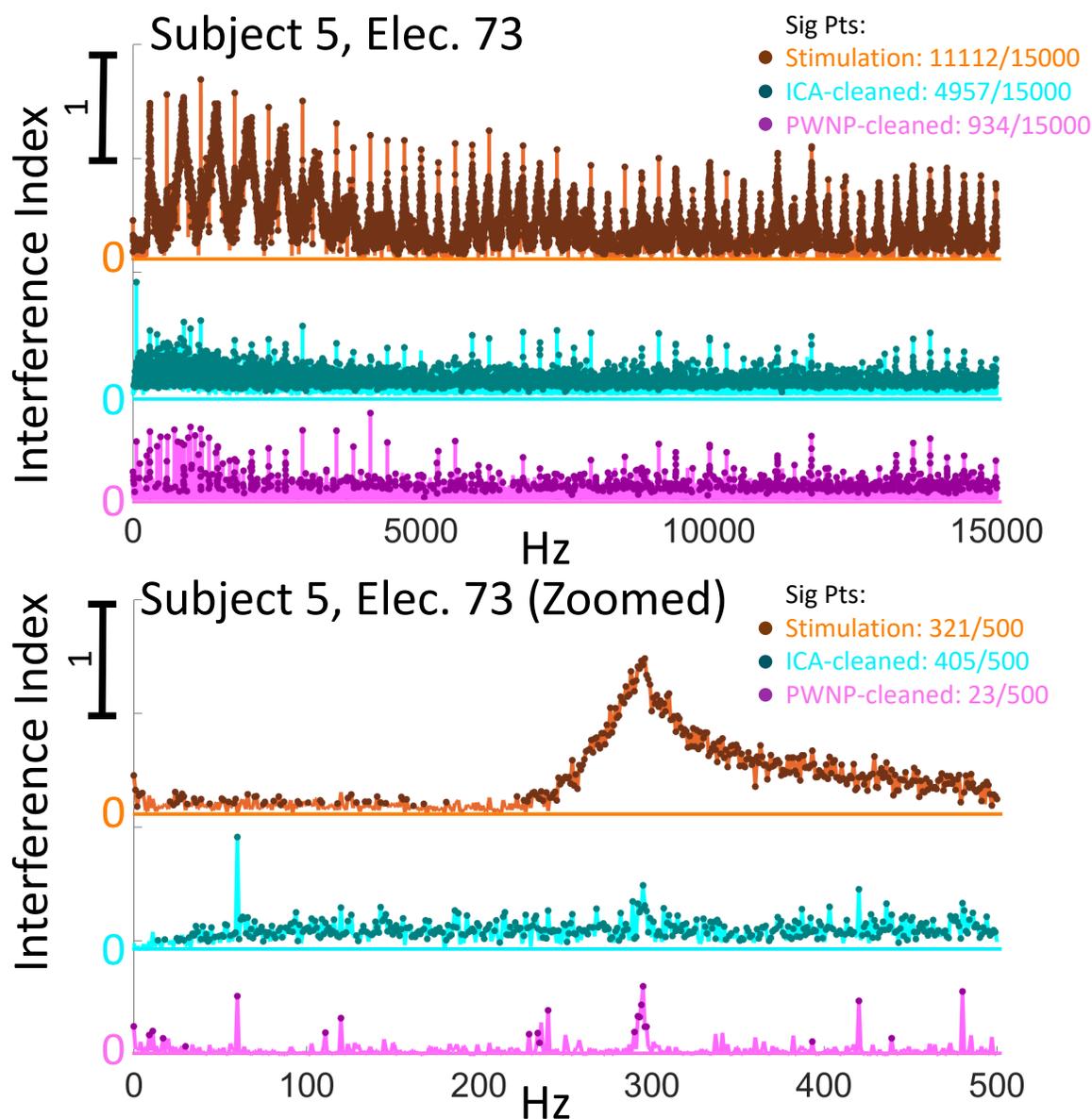
**Table 3.** Frequency-averaged interference indices,  $\bar{I}$ , and corresponding standard deviations,  $\sigma_I$ , for stimulation, PWNP-cleaned, and ICA-cleaned conditions for worst-case electrode in MEA data.

Subject 5			
	Stim	PWNP	ICA
$\bar{I}$	0.292	0.045	0.089
$\sigma_I$	0.253	0.052	0.071

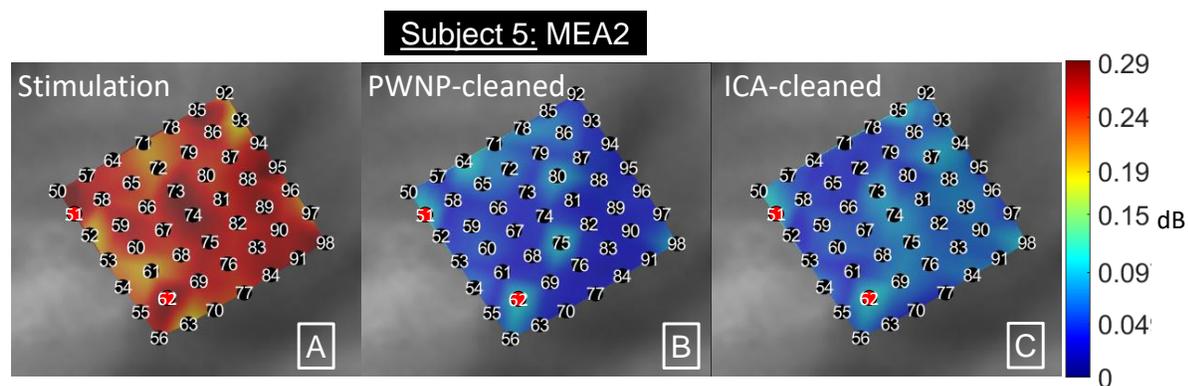
The interference index of the stimulation data exhibited peaks at the stimulation frequency (294 Hz) and its super-harmonics, similar to their PSD. Figure 12 illustrates this phenomenon for the worst-case electrode. Both artifact suppression methods reduced the artifact peaks, with PWNP generally outperforming ICA. This was evidenced by PWNP achieving lower overall interference index values, for which summary statistics are shown in Table 3. PWNP also yielded signals with fewer significant residual interference frequencies in comparison to ICA (rank-sum test,  $p < 0.01$ ). These advantages were especially evident in the local field potential range (0-500 Hz).

To visualize the effectiveness of artifact suppression methods for the whole MEA, we spatially mapped the interference indices (Figure 13). Similar to the ECoG results, the map corresponding to the stimulation data exhibited the highest values. Likewise, upon artifact suppression, these map values were significantly reduced (paired, left-tailed, signed rank test; PWNP,  $p = 8.4 \times 10^{-11}$ ; ICA,  $p = 8.4 \times 10^{-11}$ ). Consistent with our worst-case electrode analyses, we observed that PWNP outperformed ICA. Specifically, the PWNP method achieved lower interference indices across electrodes (paired, left-tailed, signed rank test,  $p = 0.000131$ ).

Figure 14 shows the results of the baseline control experiment. As with the ECoG experiments, we expect artifact suppression methods to yield small RMSE values. The average RMSE value across electrodes was  $17.4 \pm 2.6 \mu\text{V}$  for the PWNP method, which is approximately 5.4% of the pre-cleaning baseline voltage swing ( $325 \mu\text{V}$ ). The ICA method, on the other hand, yielded much larger RMSE values ( $40.9 \pm 15.5 \mu\text{V}$ ), and in turn, more significant signal distortions. Figure 14 shows time domain distortions for representative baseline signals after artifact suppression. These representative signals were taken from the electrode exhibiting the RMSE closest to the median RMSE across the grid. Similar to the ECoG experiments, the distortions introduced by PWNP were less prominent than those introduced by ICA. We also characterized these post-suppression baseline distortions in the frequency domain. Specifically, by comparing the PSDs before and after artifact suppression (signed rank test,  $p < 0.01$ ), we identified 162/15000 (1.1%) frequencies exhibiting significantly different power distributions after PWNP artifact suppression. In contrast, 1153/15000 (7.7%) frequencies were identified for ICA. To avoid potential performance bias towards the PWNP method, we performed



**Figure 12.** Interference index for the worst-case electrode from MEA2. Solid-colored lines are the interference index for the stimulation, ICA-cleaned, and PWNP-cleaned conditions. Filled-circle markers indicate frequencies where power distributions are significantly different from the baseline condition (rank-sum test,  $p < 0.01$ ). PWNP generally exhibited lower interference indices, as well as fewer frequencies with significant residual interference. (Top) Full bandwidth. (Bottom) Zoomed to local field potential bandwidth (0-500 Hz).



**Figure 13.** Interference indices spatially interpolated, color-coded and mapped to estimated MEA2 locations. (A) Interference indices for stimulation data. (B) Interference indices for PWNP-cleaned data. (C) Interference indices for ICA-cleaned data. Note that electrodes 51 and 62 were saturated, so their values were imputed to the highest in the grid to preserve the interpolation continuity.

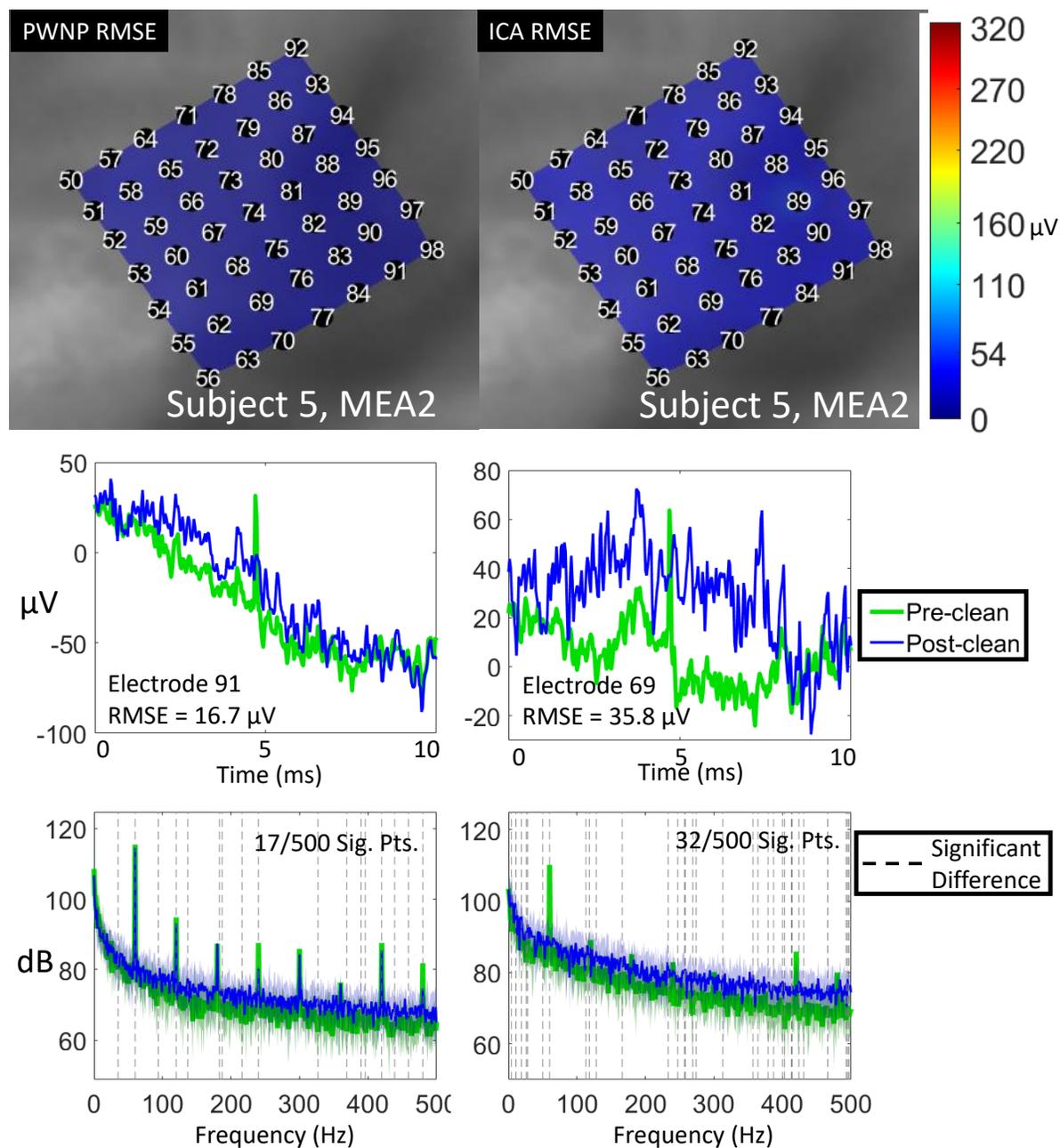
control experiments on additional baseline epochs (Appendix C), while retaining the same PWNP and ICA parameters (same procedure as ECoG baseline experiments). After averaging performances across 100 baseline epochs, we found that PWNP yielded an RMSE that accounted for only 4.9% of the baseline voltage swing. The distortions due to ICA were twice times as high, with the average RMSE value reaching 9.5% of the baseline voltage swing. These advantages were preserved in the frequency domain, where we found that PWNP on average yielded 13.0% of frequencies with significantly different power distribution compared to 28.2% of the frequencies yielded by ICA. The reader is referred to Appendix C for more information regarding these results.

### 3.4. Action Potential Recovery in MEA Stimulation Data

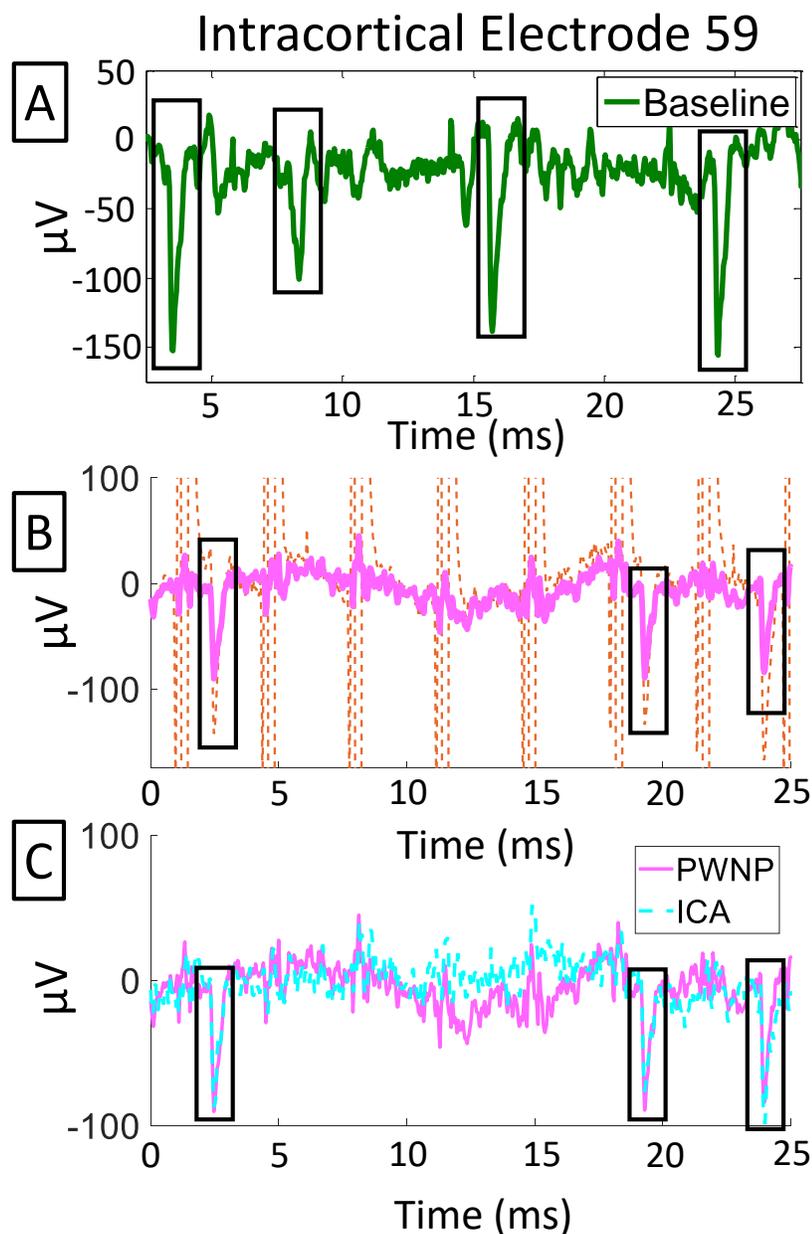
In the absence of stimulation data containing controlled behavioral tasks, we sought evidence for the PWNP method’s ability to preserve neural features in recorded data. To this end, we observed stimulation data from an electrode containing action potentials (Electrode 59) before and after artifact suppression. Across the ten stimulation epochs a total of 135 action potentials were visually identified among the stimulation artifacts. The representative time domain data segments shown in Figure 15 demonstrate the ability of PWNP and ICA methods to selectively suppress the stimulation artifacts while preserving action potentials. Both techniques had a 100% retrieval rate of the observed action potentials.

## 4. Discussion

In EEG data contaminated with narrow-band artifacts, PWNP achieved substantial artifact suppression (expressed as SIR) while preserving the underlying  $\alpha$ -band



**Figure 14.** Baseline control experiment results for PWNP (Left) and ICA (Right) methods. (Top) RMSE values spatially interpolated, color-coded, and mapped to estimated MEA2 electrode locations. Color bar range is from  $0 \mu\text{V}$  to  $325 \mu\text{V}$  (maximum absolute voltage of the baseline data). (Middle) Representative baseline time domain examples before and after artifact suppression. (Bottom) Representative frequency domain examples (mean PSD, zoomed to local field potential band, 0-500 Hz) from the same electrode before/after artifact suppression. Dashed lines indicate frequencies where the power distribution significantly differed before and after artifact suppression (signed rank test,  $p < 0.01$ ).



**Figure 15.** An example of action potentials in MEA data before and after application of PWNP and ICA methods. (A) Baseline data from an electrode exhibiting spontaneous action potentials (outlined in black). (B) Stimulation and PWNP-cleaned data from the same electrode. Stimulation artifacts reached  $\sim 3$  mV amplitudes but have been truncated to preserve scale. (C) The same data segment as in (B), comparing the PWNP-cleaned and ICA-cleaned conditions.

modulation (expressed as SNR). Specifically, PWNP effectively improved the SIR by a median of 32-34 dB, while preserving the SNR ( $|\Delta\text{SNR}| \leq 0.18$  dB). Artifact suppression as high as 44-47 dB was achieved on electrodes adjacent to the stimulation channel (C3 for Subject 1 and Cz for Subject 2). Compared to ICA, PWNP achieved superior artifact suppression results, and these differences were statistically significant for both subjects. On the other hand, ICA yielded slightly higher SNR improvements ( $|\Delta\text{SNR}| \leq 0.59$  dB), which is not surprising given ICA's ability to extract neural sources [48, 49]. However, the primary objective of artifact suppression is SIR maximization, and we only used SNR to measure whether artifact suppression compromised physiological information. If the objective is SNR maximization, we contend that artifact suppression should be followed by more appropriate SNR-enhancing methods, like supervised learning techniques [50].

Extending our analysis to broadband ECoG artifacts, we observed that PWNP retained the ability to suppress artifact features. Specifically, on worst-case electrodes, PWNP achieved a reduction of broadband artifact features, as indicated by a decrease of the frequency-averaged interference index by 0.31-0.32. Similarly, at the fundamental stimulation frequency and its super-harmonics, interference indices were reduced by as much as 1.08-1.58. Additionally, upon PWNP artifact suppression, the fraction of frequencies with significant residual artifact contamination was reduced from 108/129 (84%) to 27.5/129 (22%), averaged across Subjects 3 and 4. Our results generalize beyond the worst-case electrode, as we observed the reduction in interference indices across the whole grid. Similar to the EEG results, the electrodes in the vicinity of the stimulation channel benefited the most from PWNP. On the other hand, given that baseline data do not contain artifacts, we expect them to be unaffected by artifact suppression. Therefore, we performed baseline control experiments and demonstrated that the PWNP method did not impose severe distortions on these signals. On average, <2% of the baseline signal frequencies were significantly affected by PWNP. These spectral differences translated into 5.5% distortion in the time domain. Similar results were obtained by applying PWNP to simulated artifact data, where on average 4.5% of the frequencies had significantly different power from those of the ground truth, with the corresponding time-domain distortions of 5.6%. When these analyses were performed with ICA, the suppression results were inferior. Specifically, broadband suppression resulted in frequency-averaged interference index reductions of only 0.20-0.29 at the worst-case electrode, with reductions as high as 0.69-1.36 at the stimulation frequency and super-harmonics. ICA suppression also yielded 66.5/129 (52%) frequencies with significant residual artifact contamination, averaged across the two subjects. The ICA method also caused more distortion to the baseline data, with an average time-domain distortion of 16.5% and 43.6% of frequencies with significantly different power distributions. The ICA performance drop was also apparent in simulated artifact data, where on average 46.5% of the frequencies had significantly different power from those of the ground truth, with the corresponding time-domain distortions of 16.7%. Collectively, these results show that PWNP outperforms ICA on artifact-contaminated ECoG, and does so without imposing significant distortion onto the baseline signals.

Repeating the above analyses, we also demonstrated the ability of PWNP to suppress broadband artifacts in MEA data. For a worst-case electrode, PWNP artifact suppression achieved a reduction of the frequency-averaged interference index of 0.25. Particularly, the reductions at the stimulation frequency and super-harmonics ranged from 0.23 to 1.37. We also observed a decrease of the fraction of artifact-contaminated frequencies from 11112/15000 (74%) to 934/15000 (6%). These worst-case electrode results generalize, as we observed a significant reduction in interference indices across the whole MEA. It is worth noting that PWNP suppressed artifacts while preserving 100% of the action potentials observed in the stimulation data. For baseline control experiments, an average of 13.0% of the baseline signal frequencies were significantly affected by PWNP. These spectral differences translated into 4.9% distortion in the time domain. In contrast, ICA yielded 4957/15000 (33%) frequencies with significant residual artifacts for the worst-case electrode. ICA broadband suppression achieved a frequency-averaged interference index reduction of 0.20, with reductions of 0.16-1.17 at the stimulation frequency and super-harmonics. The baseline control experiments also indicated that ICA altered the underlying neural signals, with an average time-domain distortion 9.5%, which translated to 28.2% of frequencies with significant power distribution difference. ICA did, however preserve the action potential features, similarly to PWNP. Overall, PWNP again outperformed ICA according to multiple performance criteria.

In addition to its superior performance, PWNP is also simpler and easier to implement. Specifically, it takes advantage of stimulation artifacts being much stronger than neural signals. This enables PWNP to group these high-energy features into the top components, thereby easing the identification of the artifact subspace, as outlined in Appendix A. In many practical applications, the amplitude of stimulation artifacts can be orders of magnitude larger than those of neural sources. For example, for a typical ECoG stimulation at 3.5 mA [10, 11] and a channel impedance of 1 k $\Omega$  [51], artifacts may reach an amplitude of 3.5 V at the stimulation channel. Based on typical voltages of ECoG signals (c.f. Figure 9), it follows that artifacts may need an attenuation of >80 dB to bring their amplitude to or below the level of neural signals. This problem is exacerbated by the proximity of motor and sensory cortices (primary targets for BD-BCI applications), where the artifact attenuation due to tissue volume conduction is insufficient. For example, our prior work demonstrates that artifacts that are an order of magnitude higher than ECoG signals are found within  $\sim$ 2 cm of a stimulation channel for a range of stimulation amplitudes [18]. State-of-the-art, front-end artifact suppression techniques may alleviate this problem, but they typically provide 30-40 dB of attenuation [24, 52]. Therefore, it is likely that the assumptions set forth by PWNP remain fulfilled, even with artifact suppression at the front-end. In theory, as long as the magnitudes of stimulation artifacts exceed those of neural sources, the PWNP method is expected to work (see Appendix A). In the unlikely situation where artifacts are comparable to or weaker than neural signals, the PWNP method may fail to separate the two subspaces. In this case, however, the presence of artifacts may not be a concern

given their low amplitude.

In contrast to PWNP, ICA returns unsorted components, thus necessitating a heuristic approach to identify the components that span the artifact subspace. This problem is akin to feature selection in pattern recognition theory [53]. While there are algorithmic approaches that guarantee the optimal solution [54], they are nonetheless of combinatorial complexity. This problem is typically tackled by rank-ordering features according to certain criteria and selecting a subset of features from this list [53], as we have done here by ordering ICs according to SIR. Note, however, that this approach may be suboptimal [55], in that a combination of components exhibiting the lowest SIRs is not guaranteed to achieve the optimal artifact suppression result.

Additionally, since ICA-based artifact suppression imposed significant distortions on artifact-free baseline data, we infer that the removed ICs contained both neural and artifact features. To investigate this trade-off, we also performed a perturbation analysis on the number of artifact ICs by adding or removing individual components. Specifically, the set of artifact components was either reduced by removing components or augmented by adding putative artifact components. This analysis indicated that baseline distortions could be mitigated by discarding fewer ICs, at the expense of inferior artifact rejection. Conversely, augmenting the set artifact ICs achieved superior artifact suppression while resulting in more severe distortions. An apparent example of this was observed in the MEA data, where an attempt to improve the artifact suppression resulted in the loss of action potential features. A likely cause of these phenomena is that broadband artifacts and neural signals may not be independent of each other. Therefore, many components identified by ICA end up containing a mixture of artifact and neural features. By suppressing these ICs, some of the neural features are invariably lost. On the other hand, leaving them in fails to suppress artifact features. Conversely, PWNP is not constrained by the stringent independence requirement. Instead, it alleviates these issues by rank-ordering its components using energy and separating neural and artifact subspaces based on a single, theoretically-justified threshold  $\alpha$ , as explained in Appendix A.

Unlike the EEG data, which had a labeled behavioral task, ECoG and MEA data did not. This limitation was imposed by the nature of the ECoG and MEA data collection. Specifically, ECoG data were collected as a part of Phase II epilepsy evaluation, where it was neither justified nor practical to interfere with the clinical procedure. On the other hand, MEA data were collected as a part of a sensory mapping task [12], which solely focused on sensory responses. To address this concern, we verified that artifact suppression did not remove neural features from ECoG and MEA data by performing extensive baseline control experiments. Using ECoG data, we also generated synthetic stimulation artifacts while preserving the spatio-temporal correlations in the original signals. We then quantified the performance of artifact suppression against the ground truth. We also demonstrated that action potentials were largely unaffected by artifact suppression in the MEA data.

To demonstrate that our results generalize to real-time operation, both PWNP

and ICA artifact suppression methods should ideally be tested within a cross-validation framework. While this could be easily implemented in PWNP using standard linear algebra tools, this procedure would be prohibitively time-consuming to perform in ICA due to its reliance on a heuristic combinatorial search. Our baseline control and simulated artifact experiments (Appendix B) demonstrate some generalization capabilities of PWNP, although not in a true cross-validation manner. Therefore, our future work will test the ability of PWNP to suppress stimulation artifacts in real time. It should be noted that PWNP is amenable to efficient real-time implementation as explained below. To train the algorithm, one would collect short data segments during both baseline (stimulator off) and stimulation (stimulator on) conditions. These epochs do not need to be contiguous and they do not need to be of the same duration. From these data, the matrices  $\Sigma_B$  and  $\mathbf{H}$  can be estimated offline (following the steps in Appendix A) and saved for real-time application. In the present study, a total of 10 s, 20 s and 120 s of data was sufficient to train these parameters for the ECoG, MEA, and EEG signals, respectively. The real-time artifact suppression then reduces to matrix multiplications (Equation 1), with  $\mathbf{X}_S$  being the most recently acquired real-time data buffer. Being a purely spatial signal processing method, PWNP is independent of the buffer size. Since PWNP does not impose substantial distortions to artifact-free data, it could be added to the standard data acquisition pipeline. Alternatively, it could be synchronized with the stimulator and switched on/off accordingly.

## 5. Conclusion

In this work, we present a novel stimulation artifact suppression algorithm based on pre-whitening and null projection techniques (PWNP). We demonstrate its effectiveness in suppressing various types of stimulation artifacts across a variety of neural signals, including EEG, ECoG and MEA data. When compared to an ICA-based method, considered to be the state-of-the-art, PWNP generally demonstrated superior artifact suppression results. In conjunction with a straightforward real-time implementation, these results suggest that the PWNP algorithm is a suitable method for real-time artifact suppression in BD-BCI.

## 6. Funding

This work was funded by the National Science Foundation (Award No. 1446908, 1646275)

## 7. Conflicts of Interest

The authors declare no conflict of interests concerning the publication of this manuscript

## 8. References

- [1] A. Kübler, N. Neumann, J. Kaiser, B. Kotchoubey, T. Hinterberger, and N. P. Birbaumer, “Brain-computer communication: self-regulation of slow cortical potentials for verbal communication,” *Archives of physical medicine and rehabilitation*, vol. 82, no. 11, pp. 1533–1539, 2001.
- [2] M. Vansteensel, E. Pels, M. Bleichner, M. Branco, T. Denison, Z. Freudenburg, P. Gosselaar, S. Leinders, T. Ottens, M. Van Den Boom, *et al.*, “Fully implanted brain–computer interface in a locked-in patient with ALS,” *New Engl. J. Med.*, vol. 375, no. 21, pp. 2060–2066, 2016.
- [3] J. R. Wolpaw, R. S. Bedlack, D. J. Reda, R. J. Ringer, P. G. Banks, T. M. Vaughan, S. M. Heckman, L. M. McCane, C. S. Carmack, S. Winden, *et al.*, “Independent home use of a brain-computer interface by people with amyotrophic lateral sclerosis,” *Neurology*, vol. 91, no. 3, pp. e258–e267, 2018.
- [4] A. H. Do, P. T. Wang, C. E. King, S. N. Chun, and Z. Nenadic, “Brain-computer interface controlled robotic gait orthosis,” *Journal of neuroengineering and rehabilitation*, vol. 10, no. 1, pp. 1–9, 2013.
- [5] C. E. King, P. T. Wang, C. M. McCrimmon, C. C. Chou, A. H. Do, and Z. Nenadic, “The feasibility of a brain-computer interface functional electrical stimulation system for the restoration of overground walking after paraplegia,” *Journal of neuroengineering and rehabilitation*, vol. 12, no. 1, pp. 1–11, 2015.
- [6] A. L. Benabid, T. Costecalde, A. Eliseyev, G. Charvet, A. Verney, S. Karakas, M. Foerster, A. Lambert, B. Morinière, N. Abroug, *et al.*, “An exoskeleton controlled by an epidural wireless brain–machine interface in a tetraplegic patient: a proof-of-concept demonstration,” *The Lancet Neurology*, vol. 18, no. 12, pp. 1112–1122, 2019.
- [7] G. Pfurtscheller, G. R. Müller-Putz, J. Pfurtscheller, and R. Rupp, “Eeg-based asynchronous bci controls functional electrical stimulation in a tetraplegic patient,” *EURASIP Journal on Advances in Signal Processing*, vol. 2005, no. 19, pp. 1–4, 2005.
- [8] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, “Neuronal ensemble control of prosthetic devices by a human with tetraplegia,” *Nature*, vol. 442, no. 7099, pp. 164–171, 2006.
- [9] W. Wang, J. Collinger, A. Degenhart, E. Tyler-Kabara, A. Schwartz, D. Moran, D. Weber, B. Wodlinger, R. Vinjamuri, R. Ashmore, *et al.*, “An electrocorticographic brain interface in an individual with tetraplegia,” *PloS one*, vol. 8, no. 2, p. e55344, 2013.
- [10] S. Hiremath, E. Tyler-Kabara, J. Wheeler, D. Moran, R. Gaunt, J. Collinger, S. Foldes, D. Weber, W. Chen, M. Boninger, *et al.*, “Human perception of electrical stimulation on the surface of somatosensory cortex,” *PloS one*, vol. 12, no. 5, p. e0176020, 2017.
- [11] B. Lee, D. Kramer, M. Salas, S. Kellis, D. Brown, T. Dobreva, C. Klaes, C. Heck, C. Liu, and R. Andersen, “Engineering artificial somatosensation through cortical stimulation in humans,” *Frontiers in systems neuroscience*, vol. 12, 2018.
- [12] M. A. Salas, L. Bashford, S. Kellis, M. Jafari, H. Jo, D. Kramer, K. Shanfield, K. Pejisa, B. Lee, C. Y. Liu, *et al.*, “Proprioceptive and cutaneous sensations in humans elicited by intracortical microstimulation,” *Elife*, vol. 7, p. e32904, 2018.
- [13] S. N. Flesher, J. L. Collinger, S. T. Foldes, J. M. Weiss, J. E. Downey, E. C. Tyler-Kabara, S. J. Bensmaia, A. B. Schwartz, M. L. Boninger, and R. A. Gaunt, “Intracortical microstimulation of human somatosensory cortex,” *Science translational medicine*, vol. 8, no. 361, pp. 361ra141–361ra141, 2016.
- [14] C. Hughes, A. Herrera, R. Gaunt, and J. Collinger, “Bidirectional brain-computer interfaces,” in *Handbook of clinical neurology*, vol. 168, pp. 163–181, Elsevier, 2020.
- [15] S. Flesher, J. Downey, J. Collinger, S. Foldes, J. Weiss, E. Tyler-Kabara, S. Bensmaia, A. Schwartz, M. Boninger, and R. Gaunt, “Intracortical microstimulation as a feedback source for brain-computer interface users,” in *Brain-Computer Interface Research*, pp. 43–54, Springer, 2017.
- [16] J. Lim, P. Wang, A. Bidhendi, O. Arasteh, S. Shaw, M. Armacost, H. Gong, C. Liu, A. Do,

- P. Heydari, and Z. Nenadic, "Characterization of stimulation artifact behavior in simultaneous electrocorticography grid stimulation and recording," in *Proc. 40th Ann. Int. Conf. IEEE EMBS*, pp. 4748–4751, 2018.
- [17] A. Zhou, B. Johnson, and R. Muller, "Toward true closed-loop neuromodulation: artifact-free recording during stimulation," *Current opinion in neurobiology*, vol. 50, pp. 119–127, 2018.
- [18] J. Lim, P. T. Wang, S. J. Shaw, H. Gong, M. Armacost, C. Y. Liu, A. H. Do, P. Heydari, and Z. Nenadic, "Artifact propagation in subdural cortical electrostimulation: Characterization and modeling," *Frontiers in Neuroscience*, vol. 16, 2022.
- [19] W. Jiang, V. Hokyhkyan, H. Chandrakumar, V. Karkare, and D. Marković, "A $\pm$ 50-mv linear-input-range vco-based neural-recording front-end with digital nonlinearity correction," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 1, pp. 173–184, 2016.
- [20] H. Chandrakumar and D. Marković, "A high dynamic-range neural recording chopper amplifier for simultaneous neural recording and stimulation," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 3, pp. 645–656, 2017.
- [21] K. Limnusun, H. Lu, H. J. Chiel, and P. Mohseni, "Real-time stimulus artifact rejection via template subtraction," *IEEE transactions on biomedical circuits and systems*, vol. 8, no. 3, pp. 391–400, 2013.
- [22] A. Mendrela, J. Cho, J. Fredenburg, V. Nagaraj, T. Netoff, M. Flynn, and E. Yoon, "A bidirectional neural interface circuit with active stimulation artifact cancellation and cross-channel common-mode noise suppression," *IEEE J. Solid-St. Circ.*, vol. 51, no. 4, pp. 955–965, 2016.
- [23] J. Lim, P. Wang, H. Pu, C. Liu, S. Kellis, R. Andersen, P. Heydari, A. Do, and Z. Nenadic, "Dipole cancellation as an artifact suppression technique in simultaneous electrocorticography stimulation and recording," in *Proc. 9th Int. IEEE/EMBS Conf. Neural Eng.*, pp. 725–729, 2019.
- [24] H. Pu, J. Lim, S. Kellis, C. Y. Liu, R. A. Andersen, A. H. Do, P. Heydari, and Z. Nenadic, "Optimal artifact suppression in simultaneous electrocorticography stimulation and recording for bi-directional brain-computer interface applications," *Journal of neural engineering*, vol. 17, no. 2, p. 026038, 2020.
- [25] E. B. Montgomery Jr, J. T. Gale, and H. Huang, "Methods for isolating extracellular action potentials and removing stimulus artifacts from microelectrode recordings of neurons requiring minimal operator intervention," *Journal of neuroscience methods*, vol. 144, no. 1, pp. 107–125, 2005.
- [26] J. M. Weiss, S. N. Flesher, R. Franklin, J. L. Collinger, and R. A. Gaunt, "Artifact-free recordings in human bidirectional brain–computer interfaces," *Journal of Neural Engineering*, vol. 16, no. 1, p. 016002, 2018.
- [27] A. Zhou, S. R. Santacruz, B. C. Johnson, G. Alexandrov, A. Moin, F. L. Burghardt, J. M. Rabaey, J. M. Carmena, and R. Muller, "Wand: A 128-channel, closed-loop, wireless artifact-free neuromodulation device," *arXiv preprint arXiv:1708.00556*, 2017.
- [28] L. F. Heffer and J. B. Fallon, "A novel stimulus artifact removal technique for high-rate electrical stimulation," *Journal of neuroscience methods*, vol. 170, no. 2, pp. 277–284, 2008.
- [29] C. Waddell, J. A. Pratt, B. Porr, and S. Ewing, "Deep brain stimulation artifact removal through under-sampling and cubic-spline interpolation," in *2009 2nd International Congress on Image and Signal Processing*, pp. 1–5, IEEE, 2009.
- [30] T. Hashimoto, C. M. Elder, and J. L. Vitek, "A template subtraction method for stimulus artifact removal in high-frequency deep brain stimulation," *Journal of neuroscience methods*, vol. 113, no. 2, pp. 181–186, 2002.
- [31] T. Wichmann, "A digital averaging method for removal of stimulus artifacts in neurophysiologic experiments," *Journal of neuroscience methods*, vol. 98, no. 1, pp. 57–62, 2000.
- [32] C. Jutten and J. Herault, "Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture," *Signal processing*, vol. 24, no. 1, pp. 1–10, 1991.
- [33] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung,

- and H. H. Liu, “The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [34] J. Hofmanis, R. A. S. Ruiz, O. Caspary, R. Ranta, and V. Louis-Dorr, “Extraction of deep brain stimulation (dbs) source in seeg using emd and ica,” in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 834–837, IEEE, 2011.
- [35] T.-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski, “Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects,” *Clinical Neurophysiology*, vol. 111, no. 10, pp. 1745–1758, 2000.
- [36] T. Al-ani, F. Cazettes, S. Palfi, and J.-P. Lefaucheur, “Automatic removal of high-amplitude stimulus artefact from neuronal signal recorded in the subthalamic nucleus,” *Journal of neuroscience methods*, vol. 198, no. 1, pp. 135–146, 2011.
- [37] K. Zeng, D. Chen, G. Ouyang, L. Wang, X. Liu, and X. Li, “An eemd-ica approach to enhancing artifact rejection for noisy multivariate neural data,” *IEEE transactions on neural systems and rehabilitation engineering*, vol. 24, no. 6, pp. 630–638, 2015.
- [38] P. Wang, C. McCrimmon, P. Heydari, A. Do, and Z. Nenadic, “Subspace-based suppression of cortical stimulation artifacts,” in *Proc. 40th Ann. Int. Conf. IEEE EMBS*, pp. 2426–2429, 2018.
- [39] P. Merton and H. Morton, “Stimulation of the cerebral cortex in the intact human subject,” *Nature*, vol. 285, no. 5762, pp. 227–227, 1980.
- [40] S.-C. Wu, A. Swindlehurst, P. Wang, and Z. Nenadic, “Projection vs. prewhitening for EEG interference suppression,” *IEEE T. Bio-med. Eng.*, vol. 59, no. 5, pp. 1329–1338, 2012.
- [41] J. P. C. Da Costa, K. Liu, H. C. So, S. Schwarz, M. Haardt, and F. Römer, “Multidimensional prewhitening for enhanced signal reconstruction and parameter estimation in colored noise with kronecker correlation structure,” *Signal Processing*, vol. 93, no. 11, pp. 3209–3226, 2013.
- [42] A. Hyvarinen, “Fast and robust fixed-point algorithms for independent component analysis,” *IEEE T. Neural Networ.*, vol. 10, no. 3, pp. 626–634, 1999.
- [43] F. Cong, T. Ristaniemi, and H. Lyytinen, *Advanced signal processing on brain event-related potentials: filtering ERPs in time, frequency and space domains sequentially and simultaneously*, vol. 13. World Scientific, 2015.
- [44] D. Lehmann, “Multichannel topography of human alpha eeg fields,” *Electroencephalography and clinical neurophysiology*, vol. 31, no. 5, pp. 439–449, 1971.
- [45] S. Kay, *Fundamentals of Statistical Signal Processing. Detection Theory*. Prentice-Hall, Englewood Cliffs, New Jersey, 1998.
- [46] K. Fukunaga, *Introduction to statistical pattern recognition*. Elsevier, 2013.
- [47] S. Kaye, *Fundamentals of Statistical Signal Processing. Detection Theory*. Prentice-Hall, Upper Saddle River, New Jersey, 1989.
- [48] S. Makeig, A. Bell, T.-P. Jung, and T. J. Sejnowski, “Independent component analysis of electroencephalographic data,” *Advances in neural information processing systems*, vol. 8, 1995.
- [49] A. J. Bell and T. J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” *Neural computation*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [50] Z. Nenadic, “Information discriminant analysis: Feature extraction with an information-theoretic objective,” *IEEE T. Pattern Anal.*, vol. 29, no. 8, pp. 1394–1407, 2007.
- [51] K. A. Sillay, P. Rutecki, K. Cicora, G. Worrell, J. Drazkowski, J. J. Shih, A. D. Sharan, M. J. Morrell, J. Williams, and B. Wingeier, “Long-term measurement of impedance in chronically implanted depth and subdural electrodes during responsive neurostimulation in humans,” *Brain stimulation*, vol. 6, no. 5, pp. 718–726, 2013.
- [52] A. E. Mendrela, J. Cho, J. A. Fredenburg, V. Nagaraj, T. I. Netoff, M. P. Flynn, and E. Yoon, “A bidirectional neural interface circuit with active stimulation artifact cancellation and cross-channel common-mode noise suppression,” *IEEE Journal of Solid-State Circuits*, vol. 51, no. 4, pp. 955–965, 2016.
- [53] J. Kittler, “Feature set search algorithms,” *Pattern recognition and signal processing*, pp. 41–60,

1978.

- [54] P. M. Narendra and K. Fukunaga, “A branch and bound algorithm for feature subset selection,” *IEEE Transactions on computers*, vol. 26, no. 09, pp. 917–922, 1977.
- [55] T. M. Cover and J. M. Van Campenhout, “On the possible orderings in the measurement selection problem,” *IEEE transactions on systems, man, and cybernetics*, vol. 7, no. 9, pp. 657–661, 1977.
- [56] J. J. Ermer, J. C. Mosher, M. Huang, and R. M. Leahy, “Paired meg data set source localization using recursively applied and projected (rap) music,” *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 9, pp. 1248–1260, 2000.
- [57] M. Wax and T. Kailath, “Detection of signals by information theoretic criteria,” *IEEE Transactions on acoustics, speech, and signal processing*, vol. 33, no. 2, pp. 387–392, 1985.
- [58] M. Sun, “An efficient algorithm for computing multishell spherical volume conductor models in eeg dipole source localization,” *IEEE transactions on biomedical engineering*, vol. 44, no. 12, pp. 1243–1252, 1997.

## Appendix A. Null Projection for Artifact Suppression

During stimulation, the measurements from  $n$  sensors can be modeled as:

$$\mathbf{X}_S = \mathbf{A}_S \mathbf{S}_S + \mathbf{A}_N \mathbf{S}_N + \mathbf{N} \quad (\text{A.1})$$

where  $\mathbf{S}_S \in \mathbb{R}^{d \times t_S}$  are the time-dependent moment magnitudes of  $d$  equivalent stimulation dipoles ( $d < n$ ),  $\mathbf{S}_N \in \mathbb{R}^{s \times t_S}$  are the activities of  $s$  neural sources, and  $\mathbf{N} \in \mathbb{R}^{n \times t_S}$  is background noise [40]. The columns of  $\mathbf{A}_S \in \mathbb{R}^{n \times d}$  and  $\mathbf{A}_N \in \mathbb{R}^{n \times s}$  are the lead field vectors of the stimulation dipoles and neural sources, respectively.

If the stimulation dipoles are the strongest component in (A.1) most of the energy of  $\mathbf{X}_S$  will be contained in a  $d$ -dimensional subspace spanned by the columns of  $\mathbf{A}_S$ . This assumption holds in most practical applications, as the amplitude of artifacts caused by stimulation dipoles are typically orders of magnitude larger than those of neural sources. The stimulation artifacts can then be suppressed by null projection [40], i.e., by projecting  $\mathbf{X}_S$  to the orthogonal complement of this artifact subspace. The artifact suppression process may be hindered by strong spatial correlations present in the artifact-free signal  $\mathbf{A}_N \mathbf{S}_N + \mathbf{N}$ . These correlations may be caused by the physical proximity of individual sensors and the correlated nature of background noise.

To remove these correlations, we employ the so-called dual-condition experimental design [56], where we also collect data while the stimulator is turned off (baseline state):

$$\mathbf{X}_B = \mathbf{A}_N \mathbf{S}_N + \mathbf{N} \quad (\text{A.2})$$

To improve the SNR and accuracy of an artifact subspace estimate [41], we calculate the pre-whitening matrix from the baseline data:  $\Sigma_B^{-\frac{1}{2}} = \mathbf{V}_B \mathbf{\Lambda}_B^{-\frac{1}{2}} \mathbf{V}_B^T$ . Here,  $\Sigma_B \in \mathbb{R}^{n \times n}$  is the covariance of  $\mathbf{X}_B$ , and  $\mathbf{V}_B \in \mathbb{R}^{n \times n}$  and  $\mathbf{\Lambda}_B \in \mathbb{R}^{n \times n}$  are its eigenvector and eigenvalue matrix, respectively. Note that the pre-whitening matrix is well-defined since  $\mathbf{X}_B \in \mathbb{R}^{n \times t_S}$  is generally a full row-rank matrix ( $n \ll t_S$ ). Subsequently, we de-mean and de-correlate the stimulation data:

$$\mathbf{X}'_S = \Sigma_B^{-\frac{1}{2}} (\mathbf{X}_S - \mu_S \mathbf{1}^T) \quad (\text{A.3})$$

where  $\mu_S$  is the time-average of  $\mathbf{X}_S$ , i.e.,  $\mu_S = \frac{1}{t_s} \sum_{i=1}^{t_s} \mathbf{X}_S(i) \in \mathbb{R}^{n \times 1}$  and  $\mathbf{1} \in \mathbb{R}^{t_s \times 1}$  is a vector whose entries are all 1. Since this transformation whitens the artifact-free response in (A.1), the artifact subspace can be identified through the singular value decomposition of  $\mathbf{X}'_S$  [40]:

$$\mathbf{X}'_S = \mathbf{U}_S \boldsymbol{\Sigma}_S \mathbf{V}_S^T = [\mathbf{U}_d \quad \mathbf{U}_d^c] \begin{bmatrix} \boldsymbol{\Sigma}_d & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_d^c & \mathbf{0} \end{bmatrix} \mathbf{V}_S^T \quad (\text{A.4})$$

The artifact subspace is spanned by the columns of  $\mathbf{U}_d \in \mathbb{R}^{n \times d}$  which are the left singular vectors of  $\mathbf{X}'_S$  corresponding to its largest  $d$  singular values (the diagonal of  $\boldsymbol{\Sigma}_d$ ). The remaining  $n - d$  left singular vectors (columns of  $\mathbf{U}_d^c$ ) span the orthogonal complement of the artifact subspace. In theory, such a decomposition is possible as long as the magnitudes of stimulation dipoles exceed those of neural sources, so that the largest  $d$  singular values in (A.4) indeed correspond to stimulation artifacts. Introducing a matrix representation  $\mathbf{H} = \mathbf{U}_d^c$  and projecting data onto this subspace, i.e.,  $\mathbf{H}^T \mathbf{X}'_S$ , will result in artifact suppression, hence the name null projection. We can then reconstruct these projected data in the original space followed by ‘‘coloring’’ and restoring the mean to obtain:

$$\mathbf{X}_S^{\text{clean}} = \boldsymbol{\Sigma}_B^{\frac{1}{2}} \mathbf{H} \mathbf{H}^T \mathbf{X}'_S + \mu_S \mathbf{1}^T \quad (\text{A.5})$$

Equation (1) follows by combining (A.3) and (A.5).

In the presence of a single stimulation dipole, the theoretical dimension of the artifact subspace is  $d = 1$ . In practice,  $d$  is likely to be higher due to imperfections in the model (A.1). To accurately estimate it, we note that the non-artifact components of  $\mathbf{X}'_S$  are expected to be uncorrelated and with a unit variance. Therefore, the smallest  $n - d$  eigenvalues of its covariance matrix,  $\mathbf{S}_X$ , are expected to be  $\approx 1$ . Since  $\mathbf{X}'_S$  is a zero-mean signal, its (unbiased) covariance matrix is defined as  $\mathbf{S}_X = \frac{1}{t_s - 1} \mathbf{X}'_S (\mathbf{X}'_S)^T$ . After invoking (A.4), we have:

$$\mathbf{S}_X = \frac{1}{t_s - 1} \mathbf{U}_S \boldsymbol{\Sigma}_S \mathbf{V}_S^T (\mathbf{U}_S \boldsymbol{\Sigma}_S \mathbf{V}_S^T)^T = \frac{1}{t_s - 1} \mathbf{U}_S \boldsymbol{\Sigma}_S \boldsymbol{\Sigma}_S^T \mathbf{U}_S^T$$

where we have used the fact that  $\mathbf{V}_S$  is an orthogonal matrix ( $\mathbf{V}_S \mathbf{V}_S^T = \mathbf{I}$ ). Since  $\mathbf{U}_S$  is also an orthogonal matrix, the last equation becomes:

$$\mathbf{U}_S^T \mathbf{S}_X \mathbf{U}_S = \frac{1}{t_s - 1} \boldsymbol{\Sigma}_S \boldsymbol{\Sigma}_S^T = \frac{1}{t_s - 1} \begin{bmatrix} \boldsymbol{\Sigma}_d^2 & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{\Sigma}_d^c)^2 \end{bmatrix}$$

which represents the eigenvalue decomposition of the covariance matrix  $\mathbf{S}_X$ . Specifically:

$$\mathbf{U}_S^T \mathbf{S}_X \mathbf{U}_S = \boldsymbol{\Lambda}_S = \begin{bmatrix} \boldsymbol{\Lambda}_d & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_d^c \end{bmatrix} = \frac{1}{t_s - 1} \begin{bmatrix} \boldsymbol{\Sigma}_d^2 & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{\Sigma}_d^c)^2 \end{bmatrix}$$

where the diagonal of  $\boldsymbol{\Lambda}_d^c$  contains the eigenvalues of  $\mathbf{S}_X$  corresponding to the non-artifact components. From the last equation, it follows that  $\boldsymbol{\Sigma}_d^c = \sqrt{(t_s - 1) \boldsymbol{\Lambda}_d^c}$ , and

since the non-zero elements of  $\mathbf{\Lambda}_d^c$  are  $\approx 1$ , we expect the corresponding singular values to be  $\sigma \approx \sqrt{t_S - 1}$ . Conversely, the singular values corresponding to the artifact subspace are those that satisfy  $\sigma > \sqrt{t_S - 1}$ . To account for noise in the singular value distribution, we determined  $d$  by counting the number of singular values that satisfy  $\sigma > \alpha\sqrt{t_S - 1}$ , where  $\alpha > 1$ . More elaborate techniques for determining  $d$  based on information theoretic criteria can be found in [57].

## Appendix B. ECoG Control Experiments

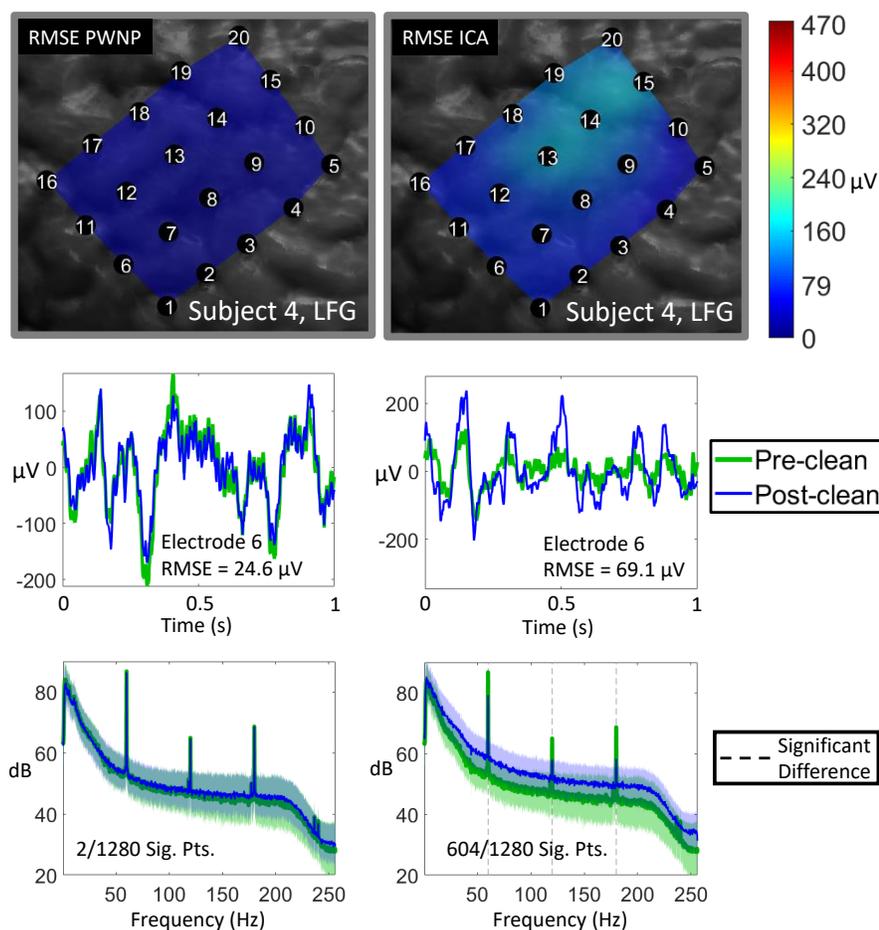
### Appendix B.1. Baseline Data

In our baseline control experiments described in Section 2.3.2, the PWNP artifact suppression method may have benefited over ICA in that the baseline data epoch had been used to calculate the pre-whitening matrix,  $\mathbf{\Sigma}_B^{-\frac{1}{2}}$ , (see Equation 1). To rule this out, we performed control experiments on additional baseline epochs, while retaining the same PWNP and ICA parameters as trained on the original data. Specifically, from the ECoG data collected from Subject 4, we segmented 100 five-second-long non-overlapping epochs of data outside of the stimulation periods.

Figure B1 summarizes the results of PWNP and ICA suppression over these 100 baseline epochs. As with the original baseline epoch, we expect the artifact suppression methods to yield small RMSE values. For PWNP, the grand average RMSE value (across epochs and channels) was  $26.3 \pm 8.1 \mu\text{V}$ , which accounted for only 5.5% of the pre-cleaning baseline voltage swing ( $475 \mu\text{V}$ ). Similar to the original baseline data (see Figure 9), the ICA method produced three times as large RMSE values ( $78.6 \pm 33.6 \mu\text{V}$ ), accounting for a 16.5% distortion of baseline signals. To appreciate these distortions in the time domain, Figure B1 also shows a representative example of baseline epoch for a representative electrode before and after artifact suppression. As before, we selected the representative electrode as the electrode exhibiting the epoch-averaged RMSE closest to the median (across the grid) epoch-averaged RMSE. Consistent with the results in Figure 9, the ICA method introduced much larger distortions in the time domain compared to the PWNP method. Finally, we characterized these distortions in the frequency domain, by comparing the epoch-averaged PSDs before and after artifact suppression (signed rank test,  $p < 0.01$ , Bonferroni corrected for multiple comparisons across frequencies and channels). Averaging across channels, we identified 24.61 out of 1280 (1.9%) frequencies exhibiting significantly different power distributions after PWNP artifact suppression and 558.06/1280 (43.6%) frequencies after ICA artifact suppression.

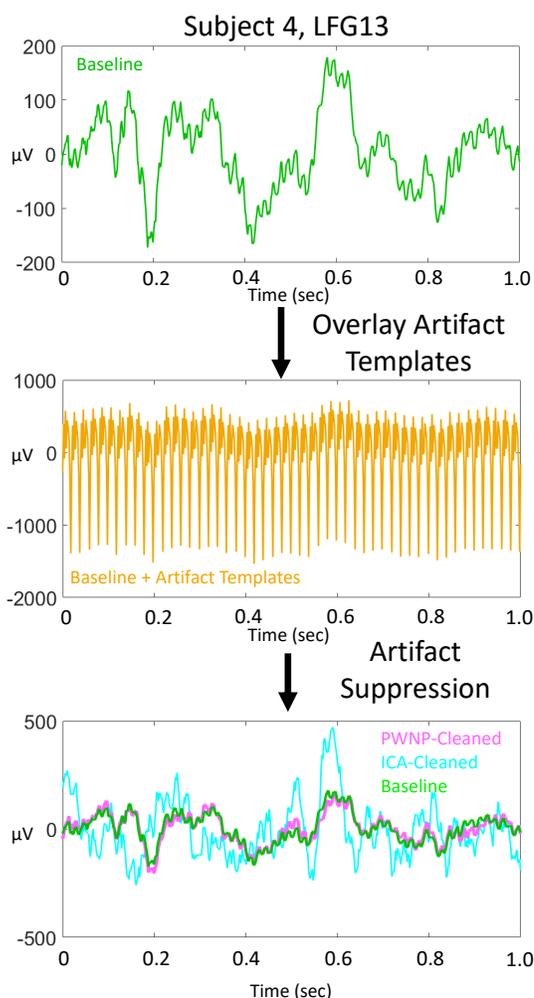
### Appendix B.2. Simulated Artifact Data

From the original stimulation epoch from Subject 4, we identified individual artifacts by performing a peak detection, constrained by the 50 Hz pulse frequency (20 ms inter-pulse



**Figure B1.** Experiment results for PWNP (Left) and ICA (Right) methods for 100 baseline epochs. (Top) RMSE values (Equation 6) averaged over 100 epochs, spatially interpolated, color-coded, and mapped to MR-CT co-registered images of the ECoG grids from Subject 4. Color bar range is from 0  $\mu\text{V}$  to 475  $\mu\text{V}$  (maximum absolute voltage across baseline epochs). (Middle) One-second segment of a representative baseline epoch before and after artifact suppression (Bottom) Epoch-averaged PSDs from the same representative electrode before/after artifact suppression. The reported numbers indicate the fraction of frequencies where the power distribution significantly differed before and after artifact suppression (signed rank test,  $p < 0.01$ , Bonferroni corrected for multiple comparisons across frequencies and channels). For clarity, the dashed lines indicate only the subset of frequencies with significant power differences across all channels.

period). To facilitate accurate estimation of artifact arrival times, we used data from the electrode LFG13, which was closest to the stimulation channel (see Figure 1) and had the largest artifacts. The arrival times were then propagated across channels, taking advantage of the fact that artifacts are phase-locked [18]. In total, for the  $\sim 5$ -second stimulation epoch, 249 artifact events were detected. Short (15.6 ms) data segments centered at each arrival time were then extracted and averaged over the 249 events to construct an artifact template for each channel. We then generated a 50 Hz train of



**Figure B2.** Generation of simulated artifacts demonstrated on the electrode closest to the stimulation channel. (Top): A one-second segment of the original uncontaminated baseline epochs from Subject 4. (Middle): The train of artifact templates overlaid onto baseline data (note the different voltage scale). (Bottom): Time-domain examples of artifact suppression results plotted alongside the ground truth baseline data.

artifact templates for each channel that superimposed it (see Figure B2) onto the same 100 baseline data epochs, as identified in the previous section. Note that this procedure preserves the spatio-temporal correlations in the original ECoG data. An alternative approach would be to generate a forward model for neuronal and artifact sources [58]; however, this model would closely match the assumptions of PWNP and ICA, which could potentially positively bias the suppression results. Finally, we applied PWNP and ICA artifact suppression methods to these simulated artifacts and performed the same comparisons using RMSE and PSDs as we did with the baseline data.

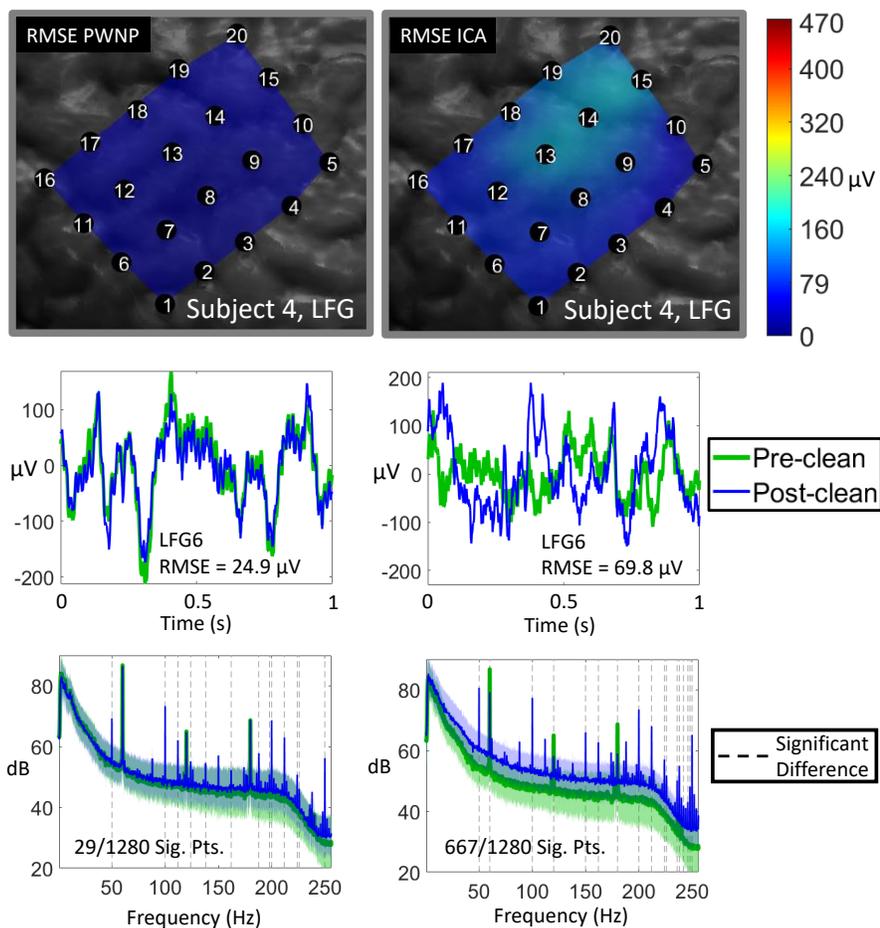
Figure B3 summarizes the artifact suppression results over the 100 simulated artifact epochs. For PWNP, the grand average RMSE value (across epochs and channels) was  $26.7 \pm 8.0 \mu\text{V}$ , which accounted for only 5.6% of the pre-cleaning baseline voltage

swing ( $475 \mu\text{V}$ ). Interestingly, these values are highly consistent with those obtained from the baseline control experiments, suggesting that the chosen threshold  $\hat{\alpha} = 1.1$  precisely delineates the artifact and neural subspaces. Similar to the baseline control experiments, the ICA method produced three times as large RMSE values ( $79.5 \pm 34.0 \mu\text{V}$ ), accounting for a 16.7% of the baseline voltage amplitude. Figure B3 shows representative examples of these distortions in the time domain for a representative electrode. As before, we selected the representative electrode as the electrode exhibiting the epoch-averaged RMSE closest to the median (across the grid) epoch-averaged RMSE. Consistent with prior results, the ICA method introduced much larger distortions in the time domain relative to the PWNP method. Finally, we characterized these distortions in the frequency domain, by comparing the epoch-averaged PSDs before and after artifact suppression (signed rank test,  $p < 0.01$ , Bonferroni corrected for multiple comparisons across frequencies and channels). After averaging across channels, we identified 57.33/1280 (4.5%) frequencies exhibiting significantly different power distributions after PWNP artifact suppression and 595.67/1280 (46.5%) frequencies after ICA artifact suppression.

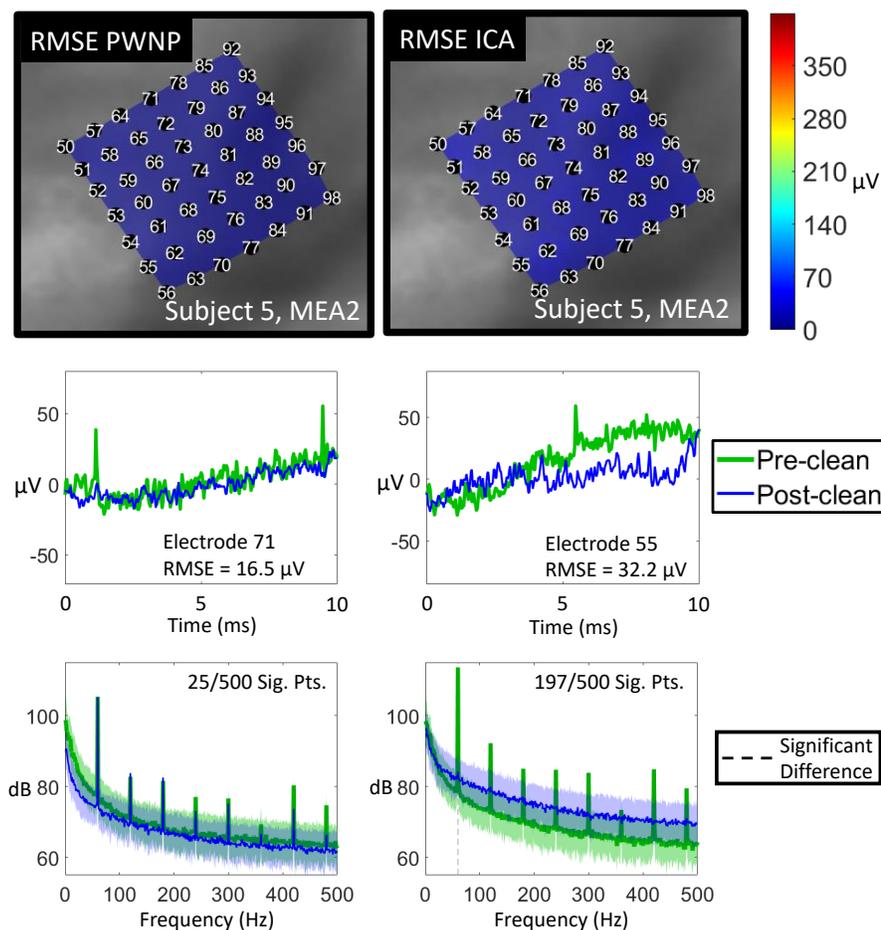
### **Appendix C. MEA Baseline Control Experiments**

Similar to the ECoG experiments, we also performed additional baseline control experiments with MEA data. Specifically, from the MEA data collected from Subject 5, we segmented 100 one-second-long, non-overlapping epochs of data outside of the stimulation periods. We then individually subjected these baseline epochs to both PWNP and ICA artifact suppression with the same originally trained parameters.

Figure C1 summarizes the results. For PWNP, the grand average RMSE value (across epochs and channels) was  $20.4 \pm 2.4 \mu\text{V}$ , which accounted for only 4.9% of the pre-cleaning baseline voltage swing ( $420 \mu\text{V}$ ). In comparison, the ICA method produced two times as large RMSE values ( $39.9 \pm 14.5 \mu\text{V}$ ), accounting for a 9.5% distortion of baseline signals. Figure C1 also shows a representative example of baseline epoch for a representative electrode before and after artifact suppression. As before, we selected the representative electrode as the electrode exhibiting the epoch-averaged RMSE closest to the median (across MEA2) epoch-averaged RMSE. Consistent with the results in Figure 14, the ICA method doubled the distortions in the time domain compared to the PWNP method. Finally, we characterized these distortions in the frequency domain, by comparing the epoch-averaged PSDs before and after artifact suppression (signed rank test,  $p < 0.01$ , Bonferroni corrected for multiple comparisons across frequencies and channels). Averaging across channels, we identified 1956.61/15000 (13.0%) frequencies exhibiting significantly different power distributions after PWNP artifact suppression and 4236.85/15000 (28.2%) frequencies after ICA artifact suppression.



**Figure B3.** Experiment results for PWNP (Left) and ICA (Right) methods for 100 simulated artifact epochs. (Top) RMSE values (Equation 6) averaged over epochs, spatially interpolated, color-coded and mapped to MR-CT co-registered images of the ECoG grids from Subject 4. Color bar range is from 0  $\mu\text{V}$  to 475  $\mu\text{V}$  (maximum absolute voltage across baseline epochs). (Middle) One-second segment of a representative baseline epoch before artifacts are overlaid and after the simulated artifacts are suppressed. (Bottom) Epoch-averaged PSDs from the same representative electrode before/after artifact suppression. The reported numbers indicate the fraction of frequencies where the power distribution significantly differed before and after artifact suppression (signed rank test,  $p < 0.01$ , Bonferroni corrected for multiple comparisons across frequencies and channels). For clarity, the dashed lines indicate only the subset of frequencies with significant power differences across all channels.



**Figure C1.** Experiment results for PWNP (Left) and ICA (Right) methods for 100 baseline epochs from MEA data. (Top) RMSE values (Equation 6) averaged over epochs, spatially interpolated, color-coded, and mapped to the estimated location of the MEA. Color bar range is from 0  $\mu\text{V}$  to 420  $\mu\text{V}$  (maximum absolute voltage across baseline epochs). (Middle) Ten-millisecond segment of a representative baseline epoch before and after artifact suppression (Bottom) Epoch-averaged PSDs from the same representative electrode before/after artifact suppression, zoomed to the local field potential band (0-500 Hz). The reported numbers indicate the fraction of frequencies where the power distribution significantly differed before and after artifact suppression (signed rank test,  $p < 0.01$ , Bonferroni corrected for multiple comparisons across frequencies and channels). For clarity, the dashed lines indicate only the subset of frequencies with significant power differences across all channels.