ERAASR: an algorithm for removing electrical stimulation artifacts from multielectrode array recordings

Daniel J O’Shea1,2 and Krishna V Shenoy1,3,4,5,6,7

1 Neurosciences Program, Stanford University, Stanford, CA 94305, United States of America
2 Department of Electrical Engineering, Stanford University, Stanford, CA 94305, United States of America
3 Department of Neurobiology, Stanford University, Stanford, CA 94305, United States of America
4 Department of Bioengineering, Stanford University, Stanford, CA 94305, United States of America
5 Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, United States of America
6 The Bio-X Program, Stanford University, Stanford, CA 94305, United States of America
7 The Stanford Neurosciences Institute, Stanford University, Stanford, CA 94305, United States of America

E-mail: djoshea@stanford.edu

Received 7 September 2017, revised 30 November 2017
Accepted for publication 21 December 2017
Published 14 February 2018

Abstract

Objective. Electrical stimulation is a widely used and effective tool in systems neuroscience, neural prosthetics, and clinical neurostimulation. However, electrical artifacts evoked by stimulation prevent the detection of spiking activity on nearby recording electrodes, which obscures the neural population response evoked by stimulation. We sought to develop a method to clean artifact-corrupted electrode signals recorded on multielectrode arrays in order to recover the underlying neural spiking activity. Approach. We created an algorithm, which performs estimation and removal of array artifacts via sequential principal components regression (ERAASR). This approach leverages the similar structure of artifact transients, but not spiking activity, across simultaneously recorded channels on the array, across pulses within a train, and across trials. The ERAASR algorithm requires no special hardware, imposes no requirements on the shape of the artifact or the multielectrode array geometry, and comprises sequential application of straightforward linear methods with intuitive parameters. The approach should be readily applicable to most datasets where stimulation does not saturate the recording amplifier. Main results. The effectiveness of the algorithm is demonstrated in macaque dorsal premotor cortex using acute linear multielectrode array recordings and single electrode stimulation. Large electrical artifacts appeared on all channels during stimulation. After application of ERAASR, the cleaned signals were quiescent on channels with no spontaneous spiking activity, whereas spontaneously active channels exhibited evoked spikes which closely resembled spontaneously occurring spiking waveforms. Significance. We hope that enabling simultaneous electrical stimulation and multielectrode array recording will help elucidate the causal links between neural activity and cognition and facilitate naturalistic sensory prostheses.

Keywords: electrical stimulation, multielectrode array, algorithm, neural populations, neural prosthetics

(Some figures may appear in colour only in the online journal)
1. Introduction

Electrical stimulation is a method of modulating neural activity that is widely used within neuroscience and neuroengineering, as well as for treatment of chronic neurological pathologies. Within neuroscience, microstimulation is used to probe the functional organization of neural circuits. Whereas electrophysiological recordings provide correlative insights that neural activity in a brain region appears related to a specific behavior, electrical stimulation can demonstrate causal contributions of brain regions to specific cognitive functions (Salzman et al. 1990). In particular, intracortical microstimulation (ICMS) has been used to identify neural networks underlying perception, attention, cognition, and movement (see Histed et al. 2010). While effective, hardware approaches may be expensive or challenging to scale to high channel count recording arrays and typically require very stable artifacts over time for proper removal. Offline approaches employ standard electrophysiology collection hardware to record neural signals and then estimate and subtract the artifact post hoc. The most common approach for artifact estimation is to simply interpolate across the artifact (O’Keeffe et al. 2001, Montgomery et al. 2005) and/or to average over repeated stimulation epochs (Hashimoto et al. 2002, Montgomery et al. 2005, Klink et al. 2017) and subtract, exploiting the larger stereotypy of the stimulus artifact relative to variable neural responses. Others have employed curve fitting approaches to estimate the artifact, exploiting the differences in the shape of artifacts and action potentials, thereby allowing for small variations in the observed artifact across trials (Wagenaar and Potter 2002, Erez et al. 2010).

These previous approaches have two major shortcomings addressed in this work. First, these approaches only recover neural signals occurring after the stimulation pulse is completed, with a typical window of several milliseconds during which signal is discarded. However, ICMS is typically delivered in trains of stimulation pulses occurring at frequencies approaching several hundred pulses per second (e.g. Churchland and Shenoy 2007, Klink et al. 2017). This failure to recover neural activity concomitant with the stimulation pulses renders a large fraction of the peri-stimulation signal unusable for spike detection and waveform discrimination. Secondly, prior work has focused on artifacts detected at a single electrode, not exploiting the similar structure of contaminating artifacts across multiple electrodes using a multielectrode array. A notable exception is the work of Mena and colleagues (Mena et al. 2011, Mena et al. 2017) and subtract, allowing for small variations in the observed artifact across trials (Wagenaar and Potter 2002, Erez et al. 2010).

These previous approaches have two major shortcomings addressed in this work. First, these approaches only recover neural signals occurring after the stimulation pulse is completed, with a typical window of several milliseconds during which signal is discarded. However, ICMS is typically delivered in trains of stimulation pulses occurring at frequencies approaching several hundred pulses per second (e.g. Churchland and Shenoy 2007, Klink et al. 2017). This failure to recover neural activity concomitant with the stimulation pulses renders a large fraction of the peri-stimulation signal unusable for spike detection and waveform discrimination. Secondly, prior work has focused on artifacts detected at a single electrode, not exploiting the similar structure of contaminating artifacts across multiple electrodes using a multielectrode array. A notable exception is the work of Mena and colleagues (Mena et al. 2011, Mena et al. 2017) and subtract, allowing for small variations in the observed artifact across trials (Wagenaar and Potter 2002, Erez et al. 2010).

These previous approaches have two major shortcomings addressed in this work. First, these approaches only recover neural signals occurring after the stimulation pulse is completed, with a typical window of several milliseconds during which signal is discarded. However, ICMS is typically delivered in trains of stimulation pulses occurring at frequencies approaching several hundred pulses per second (e.g. Churchland and Shenoy 2007, Klink et al. 2017). This failure to recover neural activity concomitant with the stimulation pulses renders a large fraction of the peri-stimulation signal unusable for spike detection and waveform discrimination. Secondly, prior work has focused on artifacts detected at a single electrode, not exploiting the similar structure of contaminating artifacts across multiple electrodes using a multielectrode array. A notable exception is the work of Mena and colleagues (Mena et al. 2011, Mena et al. 2017) and subtract, allowing for small variations in the observed artifact across trials (Wagenaar and Potter 2002, Erez et al. 2010).
and can recover the full timecourse of neural signals during and after stimulation, requiring only that the electrical artifact does not saturate the amplifier on the recording electrodes. Our algorithm exploits the similarity of artifacts across multiple electrodes, multiple pulses within a pulse train, and multiple trials of repeated stimulation. The ERAASR algorithm is structured as a sequence of cascaded, linear operations—PCA and linear regression—across each axis of the data tensor. We validate our method on neural signals collected during an arm-movement task with 24-channel linear multielectrode arrays (‘V probes’, Plexon Inc.) in macaque dorsal premotor cortex, demonstrating that spiking waveforms extracted during the stimulation period closely resemble spontaneous spiking waveforms on each electrode.

2. Methods

2.1. Subjects

Animal protocols were approved by the Stanford University Institutional Animal Care and Use Committee. The subject was one adult male macaque monkey (Macaca mulatta), monkey P. After initial training, we performed a sterile surgery during which the macaque was implanted with a head restraint and a recording cylinder (NAN Instruments) which was located over left, caudal, dorsal premotor cortex (PMd). The cylinder was placed normal to the skull and secured with methyl methacrylate. A thin layer of methyl was also deposited atop the intact, exposed skull within the chamber. Before stimulation sessions began, a small craniotomy (3 mm diameter) was made under ketamine/xylazine anesthesia, targeting an area in PMd which responded during movements and palpation of the upper arm (17 mm anterior to interaural stereotaxic zero).

2.2. Reaching task

For the purposes of this study, we engaged the monkey in a reaching task in order to drive task-related neural activity in premotor cortex where our electrodes were located. We trained the monkey to use his right hand to grasp and translate a custom 3D printed handle (Shapeways Inc.) attached to a haptic feedback device (Delta3, Force Dimension Inc.). The other arm was comfortably restrained at the monkey’s side. The monkey was trained to perform a delayed reaching task by moving the haptic device cursor towards green rectangular targets displayed on the screen. Successful completion of each movement triggered a juice reward.

2.3. Electrophysiology and stimulation configuration

At the start of each experimental session, both a stimulation electrode and a recording probe were secured to two independently controllable, motorized micromanipulators (NAN Instruments). Both probes were lowered simultaneously through blunt, non-penetrating guide tubes into dorsal premotor cortex at 3 μm s⁻¹ to an approximate depth of 2 mm (figure 1(a)). The stimulation electrode was a single tungsten microelectrode with approximately 100 kΩ impedance at 1 kHz (Part #UEWLG-SECN1E, Frederick Haer Company). The recording probe was a linear electrode array consisting of 24 contact sites located at 100 μm spacing along the length of the shank (V-probe PLX-VP-24-15ED-100-SE-100-25(640)-CT-500, Plexon Inc.). The recording probe penetration site was located at an approximate distance of 0.75–1.5 mm from the stimulation site. As best possible, the stimulating electrode was inserted at the same location and to the same depth, whereas we changed the location of the recording electrode for each recording session.

The stimulation was performed using a StimPulse electrical stimulation system used as a combined function generator and isolated current source (Frederick Haer Company). The electrical stimulation current flowed through the brain via the electrode to a ‘ground screw’ located at the posterior pole of the implant whose tip was in contact with the dura below the skull. Microstimulation trains consisted of 20 biphasic pulses delivered at 333 Hz for 57 ms (150 μs cathodic, 100 μs pause, 150 μs anodic). Stimulation amplitude was varied between 5–40 μA. Stimulation delivered at the start of most experimental sessions, while the arm was in a passive, resting position, triggered brief movements of the contralateral upper arm and shoulder at thresholds typically >140 μA. Stimulation onset was triggered via TTL pulse delivered by the task engine (Simulink Real-time Target, The Mathworks; NIDAQ digital to analog card, National Instruments). Stimulation was delivered on 20–40% of trials. Stimulation amplitude was fixed within each block of trials, where we collected at least 8–10 repetitions of each stimulation trial type for each block. Within each session, we typically began with blocks of lower amplitude stimulation before proceeding to higher amplitude stimulation.

The recording probe was connected to a 3-headed switching headstage, a component of the Blackrock StimSwitch system (Blackrock Microsystems). The probe connects to one bank of inputs, control lines from the StimSwitch control box connect to the second bank of inputs, and outgoing voltage signals leave via the third bank of outputs. During these experiments, this component was configured to act as a unity gain buffer (head stage) which simply relays the voltage signals to a shielded ribbon cable (Samtec Inc.) to the amplifier (Blackrock Microsystems). Each electrode is differentially amplified relative to a common reference line in the V-probe itself, which is also shorted to surrounding guide tube for better noise rejection.

Broadband voltages were recorded from all 24 electrode channels of the recording probe. We also recorded from the stimulation electrode during initial penetration to verify that we could see nearby neurons on the electrode. We then disconnected the stimulation electrode from the recording system and connected it to the StimPulse stimulator. Broadband signals were filtered at the amplifier (0.3 Hz 1 pole high-pass filter, 7.5 kHz 3 pole low-pass filter), digitized to 16 bit resolution over ±8.196 mV (resolution = 0.25 μV), and sampled at 30 kHz.
During experiments, we recorded any units or multi-units that could be isolated on the multielectrode array, without regard for their responsiveness or modulation by the task. Before analysis, we applied a screening procedure which looked only at non-stimulated trials in order to remove neurons that were very unreliable or very weakly modulated by the task. Briefly, we filtered non-stimulated trials spike trains with a 30 ms Gaussian window, then aligned trials separately from 100 ms pre-target onset to 70 ms post go cue, and then from 300 ms pre-movement to 600 ms post, then averaged within groups of trials with the same reach target. We defined each unit’s ‘signal’ to be the range of firing rates over time and conditions, and ‘noise’ to be the maximum standard error of the mean firing rate. We included units where the ratio of signal to noise exceeded at least 2, a total of 138 units in monkey P. Note that this selection process looked only at non-stimulation trials.

2.5. Artifact amplitude model

The amplitude of the electrical artifact was measured on each recording channel as the peak to peak voltage concomitant with the first stimulus pulse, averaged across trials. We then devised a model relating spatial location to artifact size that allowed us to pinpoint the location of the recording probe in each session relative to the stimulation source. Amplitude measurements were used to fit a model relating artifact amplitude to the distance of each site to the stimulation source. This model jointly optimizes parameters defining the distance-dependent amplitude scaling and the location of the recording probe in each session (relative to the stimulation source). We define $x_s$ as the distance measured laterally from the probe to the closest point on the stimulating electrode; we define $y_s$ as the vertical distance along the penetration path from the most superficial electrode on the probe to the tip of the stimulating electrode. The location of each electrode $e$ in each session $s$ relative to the probe location $(x, y)_s$ is given by

$$ (x, y)_{e,s} = (x_s, y_s + e\Delta) \quad (1) $$

where $\Delta = 100 \mu m$ spacing between sites. This contact site is located at a distance

$$ d_{e,s} = \sqrt{x_{e,s}^2 + y_{e,s}^2} \quad (2) $$

from the stimulation source which lies at the origin. We then assume that the amplitude of the artifact on each electrode for a specific stimulation amplitude $I_s$, falls off as the reciprocal of this distance.

$$ V_{e,s,a} = \frac{v_0 + I_s}{d_{e,s}} \quad (3) $$

where $v_0$ has units V ($\mu A$ mm)$^{-1}$. The parameters of this model are this single scaling parameter $v_0$ as well as a pair of location coordinates for the probe on each session. We then fit this model using constrained non-linear curve fitting (lsqcurvefit in Matlab) to minimize the squared error between empirical and modeled artifact amplitude across electrodes, sessions, and amplitudes. Confidence intervals of the parameters were obtained using the Jacobian of the fitted model.

We note that ERAASR makes no assumptions regarding the artifact amplitudes across electrode channels. This model is fit solely to demonstrate basic features of the evoked artifacts in our datasets and to infer the location of the recording probes.
2.6. Synthetic dataset generation

We generated synthetic datasets on which to test the performance of ERAASR when evoked spiking activity was highly reliable relative to the electrical stimulus. We applied ERAASR to real datasets and extracted the estimated artifacts for an individual channel. We then added a known, representative spiking waveform on top of this artifact at specific times during stimulation. These spikes were added to a subset of pulses at offsets drawn from a normal distribution with a fixed mean and specified standard deviation or jitter. The proportion of pulses with synthetic spikes and the jitter were jointly varied systematically. White noise was then added to the resultant traces to match that synthetic spikes and the jitter were jointly varied systematically.

2.7. Reduced channel count comparison

To probe the dependence of ERAASR performance on channel count, we applied ERAASR to subsets of channels selected from the original datasets. We first identified ‘spike-free’ channels that had no spontaneous or evoked spiking activity, which were typically very superficial. For a given channel count, we randomly selected groups of channels of that size from the original 24, ensuring that at least one of the chosen channels was spike-free. For each spike-free channel, we computed the RMS voltage of the channel post cleaning and high-pass filtering at 250 Hz and thresholding at \(-5\times\) RMS voltage to extract spiking waveforms. The mean recovered waveform was then compared to the true spiking waveform filtered by the same high-pass filter.

3. The ERAASR algorithm

We describe the operation of the ERAASR algorithm as applied to a representative example dataset recorded in macaque PMd. Pseudocode is provided in algorithm 1.

3.1. Precise temporal alignment and tensor construction

We began with the raw, unfiltered voltage traces for each stimulation trial, collected simultaneously on the 24 channels, sampled at 30 kHz. We first corrected the small amount of jitter between triggering the stimulator to the beginning of the stimulation train, typically on the order of a few milliseconds. We began by precisely aligning the stimulation trials to each other in time. This process was performed separately for each group of trials sharing a common stimulation amplitude. We extracted the raw voltage data from one electrode channel in a 60 ms window surrounding the triggering event, high-pass filtered the signal with a fourth-order, 250 Hz corner frequency Butterworth filter, to remove slow drift. We then detected the onset of the artifact due to the first biphasic pulse on each trial using an appropriately low negative threshold, set to avoid spiking activity but reliably detect artifacts at the lowest stimulation amplitude. We then realigned the voltage data for each trial to this threshold crossing event. An outlier detection procedure was performed to detect rare instances of stimulator failure, including trials where 19 or 21 pulses were delivered instead of 20. Outlier detection was performed by projecting the realigned, unfiltered voltage traces onto the first 10 principal components of that group, and detecting trials where the absolute value of the z-scored scores in any of the first 10 PCs exceeded a value of 5. Outlier trials (typically <2% of trials) were excluded from subsequent analysis. The traces were then up-sampled 10 \(\times\) to 300 kHz using spline interpolation and the set of trials were precisely aligned in time. We manually verified the results of this procedure to ensure successful artifact removal and accurate alignment of the artifact pulse trains. The aligned traces are then resampled at the original 30 kHz sampling rate.

We constructed a data tensor for each group of trials sharing a common stimulation amplitude and extract a 60 ms window of the trial-aligned broadband voltage data across the trials, yielding a data tensor \(X_{\text{raw}}\) with size \(C\) (number of channels) \(\times T\) (number of timepoints per pulse) \(\times P\) (number of pulses) \(\times R\) (number of trials). In our datasets \(C = 24\) channels, \(T = 90\) timepoints per pulse (3 ms at 30 kHz sampling), \(P = 20\) pulses, and \(R\) was typically on the order of 100–200 trials. We lightly high-pass filtered the raw signals in time with a fourth-order, 10 Hz corner frequency Butterworth filter, which removed slow drifts from the traces and makes subsequent processing more robust. We first attempted to clean each channel by exploiting the similarities of the artifact across simultaneous channels (figure 2(a)).

3.2. Removal of shared structure over channels

We then unfolded the data tensor into an \(\text{RTP} \times C\) matrix \(M\) (figure 2(b)), and performed principal components analysis. This allows us to re-express each channel’s response vector \(\text{RTP} \times 1\) as a linear combination of principal components, each \(\text{RTP} \times 1\). As the artifact waveform is very large relative to the interesting spiking activity and shared across all channels, one would expect the top few principal components to capture preferentially shared variance due to the artifact. Empirically, we determined that the top \(K_C = 4\) principal components (PCs) captured much of the artifact shape (figure 2(c)). We then reconstructed each channel’s response omitting the contribution from the artifact. The simplest method to accomplish this would be to reconstruct the channel’s response by using a linear combination of all PCs except the first \(K_C\). However, in some cases, especially when lower currents were used and the magnitude of the artifact relative to the spiking activity was smaller, we observed that some of the top PCs
would begin to incorporate a small amount of spiking activity from individual channels. By reconstructing those channels from the remaining PCs, the spiking activity itself would be partially distorted because a fractional portion of those spikes would be subtracted along with the artifact. We addressed this issue by again exploiting the locality of spiking activity, by assuming that spiking activity from any one channel would not be present on other channels except the immediately adjacent channels. However, the artifact is shared on a much larger spatial scale and can be separated from the spiking activity. Therefore, for each channel \( c \), we used the following procedure: reconstruct the top \( K_c \) PCs using all channels except \( c \) and its immediately adjacent neighbors \( c - 1 \) and \( c + 1 \), that is, using a modified version of the loading weights \( w^C_{(k)} \) for the \( k \)th principal component in which we set the weights for \( c, c - 1, c + 1 \) to 0. More generally, we define parameter \( \lambda_c \) that dictates the number of adjacent channels excluded from the reconstruction. We refer to this vector of modified loading weights as \( w^C_{(k,c)} \).

\[
\left( w^C_{(k,c)} \right) = \begin{cases} 0 & \text{if} \ c - \lambda_c \leq i \leq c + \lambda_c \\ \left( w^C_{(k)} \right)_i & \text{else} \end{cases}
\]  

Using these loading weights, we reconstruct the matrix of the top \( K_c \) artifact-capturing PCs, where each column is given by

\[
(A_c)_{(k)} = M^C w^C_{(k,c)} \quad k = 1, \ldots, K_c; \ c = 1, \ldots, C.
\]  

We then regress the channel’s response vector against these artifact components and subtract this reconstructed artifact from the channel’s response:

\[
M^C = M^C - (A^T_c A_c)^{-1} (A^T_c) M^C.
\]

This cleaning across channels captured (figure 2(d)) and removed (figure 2(e)) much of the artifact from each pulse. However, when examining the responses on an individual channel across each of the 20 pulses superimposed, we observed consistent structure in the traces (figure 2(f)).

3.3. Removal of shared structure over pulses

This structure indicated that there remained artifact structure that is unique to individual channels but shared among the pulses in each trial. To address this, we rearrange \( M^C \) into a new matrix \( M^P \) with size \( TRC \times P \), a set of responses to each individual pulse, concatenated over trials and channels (figure 3(a)). We determined that the top \( K_p = 2 \) principal components (PCs) captured much of the artifact shape. Consequently, we used the same PC regression procedure to estimate and subtract the artifact on each pulse, omitting the
contribution of the pulse under consideration and optionally adjacent pulses within $\lambda_p$. For our dataset, $\lambda_p = 0$ was used. If $w_{(k,p)}^R$ is a vector of $P$ loading weights describing the $k$th principal component, and for each pulse $p$, $w_{(k,p)}^P$ is given by:

$$
(w_{(k,p)}^P)_i = \begin{cases} 0 & \text{if } p - \lambda_p \leq i \leq p + \lambda_p \\ (w_{(k,p)}^R)_i & \text{else} \end{cases}
$$

For each pulse, we reconstructed the matrix of the top $K_P$ artifact-capturing PCs, whose columns are given by:

$$(A_p)_{(k)} = M_p^R w_{(k,p)}^R, \quad k = 1, \ldots, K_P; \quad p = 1, \ldots, P. \quad (8)$$

We then regressed the channel’s response vector against these artifact components and subtract this reconstructed artifact from the channel’s response:

$$
\hat{M}_p^R = M_p^R - (A_p^T A_p)^{-1} (A_p^T M_p^R).
$$

This cleaning procedure over pulses captured much of the common artifact structure observed within each train (figure 3(b)). It is also possible to perform this cleaning over pulses separately for each channel, following an approach like that described below for cleaning over trials.

We next examined the set of cleaned traces for individual channels and pulses in the train, over the full set of trials. We observed that there remained common artifactual structure in these traces, which suggests that there is artifact structure unique to that channel and pulse number but shared among many trials (figure 3(c)). In most cases, this shared structure was similar among nearby trials but exhibited a slow drift in the shape of the artifacts over the experimental session.

### 3.4. Removal of shared structure across trials

To remove these residual artifacts, we again employed a principal components regression approach, exploiting the shared structure of the artifact over multiple trials. We performed this operation separately for each channel. We rearranged the cleaned $M_p^R$ into a set of $C$ data matrices $M_{cR}^R$ for $c \in \{1, \ldots, C\}$, each with size $T \times P \times R$ (figure 4(a)). We determined that the top $K_R = 4$ PCs captured much of the artifact shape over trials. If $w_{(c,k)}^R$ is a vector of $R$ loading weights describing the $k$th principal component for channel $c$, we defined:

$$(w_{(c,k)}^R)_i = \begin{cases} 0 & \text{if } r - \lambda_R \leq i \leq r + \lambda_R \\ (w_{(c,k)}^R)_i & \text{else} \end{cases} \quad (9)$$

where $\lambda_R$ adjacent trials were excluded from the reconstruction. For our dataset, $\lambda_R = 0$ was used.

For each trial, we reconstructed the matrix of the top $K_R$ artifact-capturing PCs, whose columns are given by:

$$(A_{(c,r)})_{(k)} = (M_{cR}^R) (w_{(c,k)}^R), \quad k = 1, \ldots, K_R; \quad r = 1, \ldots, R; \quad c = 1, \ldots, C. \quad (10)$$

We then regressed the trial’s response vector against these artifact components and subtract this reconstructed artifact from the trial’s response:

$$
\hat{M}_{(c,r)}^R = M_{(c,r)}^R - (A_{(c,r)}^T A_{(c,r)})^{-1} (A_{(c,r)}^T M_{(c,r)}^R).
$$

We then rearranged the set of cleaned $M_{cR}^R$ back into a $C \times T \times P \times R$ tensor $X_{\text{cleaned}}$. This cleaning procedure captured much of the residual artifact (figure 4(c)). The cleaned signals in $X_{\text{cleaned}}$ did not display features obviously indicative of residual stimulation artifact, but did on some channels contain readily detectable spiking activity (figure 4(c), bottom panel). We then spliced the cleaned
data back into the full trials, adding appropriate offsets to preserve continuity at the first and last timepoints of the inserted segment.

Algorithm 1. ERAASR algorithm pseudocode.

Input: Raw data tensor \(X_{\text{raw}}(c,t,p,r)\) with size \(C\) channels by \(T\) timepoints per pulse by \(P\) pulses by \(R\) trials

Output: Cleaned data tensor \(X_{\text{cleaned}}(c,t,p,r)\) with same size

Parameters: \(K_C, K_P, K_R\) - number of principal components to describe artifact structure over channels, pulses, trials \(\lambda_C, \lambda_P, \lambda_R\) - number of adjacent channels, pulses, trials excluded as regressors during artifact reconstruction \(\beta_P, \beta_C \in \{\text{false, true}\}\) - Whether to perform cleaning over pulses separately on each channel. Below we assume \(\beta_P\) is false and \(\beta_C\) is true for clarity.

\[
X \leftarrow X_{\text{raw}}
\]

Clean across channels

Unfold \(X\) into \(M^C\) with size \(RPT\) by \(C\).

\[
W^C \leftarrow \text{projection along leading } K_C \text{ principal components of } M^C
\]

for \(c \in 1, \ldots, C\) do

\[
A_c \leftarrow M^C W^C_c \text{ with weights zeroed for channel } c \text{ and channels within } \lambda_C
\]

\[
X(c,:) \leftarrow X(c,:) - \text{projection of } X(c,:) \text{ onto } A_c
\]

end

Clean across pulses

Unfold \(X\) into \(M^P\) with size \(TRC\) by \(P\).

\[
W^P \leftarrow \text{projection along leading } K_P \text{ principal components of } M^P
\]

for \(p \in 1, \ldots, P\) do

\[
A_p \leftarrow M^P W^P_p \text{ with weights zeroed for pulse } p \text{ and pulses within } \lambda_P
\]

\[
X(:,p,:) \leftarrow X(:,p,:) - \text{projection of } X(:,p,:) \text{ onto } A_p
\]

end

Clean across trials, separately for each channel

for \(c \in 1, \ldots, C\) do

Unfold \(X(c,:)\) into \(M^R_c\) with size \(TP\) by \(R\).

\[
W^R_c \leftarrow \text{projection along leading } K_R \text{ principal components of } M^R_c
\]

for \(r \in 1, \ldots, R\) do

\[
A_{c,r} \leftarrow M^R_c W^R_{c,r} \text{ with weights zeroed for trial } r \text{ and trials within } \lambda_R
\]

\[
X(c,:,r) \leftarrow X(c,:,r) - \text{projection of } X(c,:,r) \text{ onto } A_{c,r}
\]

end

end

return \(X_{\text{cleaned}} \leftarrow X\)

3.5. Post-stimulation transient cleaning

Following stimulation offset, a slower transient was observed in all channels. These transients were highly similar among trials. Consequently, we employed a similar principal components regression procedure to remove these slow transients. We performed the following procedure for each stimulation amplitude separately. We began by aligning the trials to stimulation offset and extracting voltage data in a 30 ms time window beginning after stimulation end, yielding a tensor \(X_{\text{post}}\) with size \(R\) by \(C\). We then rearrange this tensor into a set of \(C\) data matrices \(M^C_{\text{post}}\) for \(c \in \{1, \ldots, C\}\), each with size \(T \times R\). We determined that much of the transients’ temporal structure was present in each trial for a given channel and that the top \(K_{\text{post}} = 2\) principal components (PCs) of these \(M^C_{\text{post}}\) matrices captured much of the transient shape over trials. Consequently, we repeated the same principal components regression approach described in equations (10) to (12) to clean the post-stimulation transients over trials. We then rearranged the set of cleaned \(M^C_{\text{post}}\) back into a \(R\) by \(C\) tensor \(X_{\text{post-cleaned}}\). The signals in \(X_{\text{post-cleaned}}\) were then reinserted into the full voltage traces for each stimulation trial.

3.6. Spike thresholding and sorting

The cleaned voltage traces, along with the original voltage traces on non-stimulated trials were then high-pass filtered using a fourth-order 250 Hz corner frequency Butterworth filter, as is typically done online before spike detection. We then thresholded each signal at −5x RMS voltage, using a greedy procedure in which the spike threshold crossings were detected in order of size (largest to smallest), and no further threshold crossings were considered within a lockout period extending 0.3 ms prior to and 1.0 ms after the current threshold crossing. We then hand-sorted the spiking waveforms using MKsort (Matthew Kaufman and Ripple Inc., https://github.com/ripple-neuro/mksort).

4. Results

We delivered high-frequency (333 Hz) biphasic electrical stimulation in macaque dorsal premotor cortex through a tungsten microelectrode, while simultaneously recording neural spiking activity using a 24-channel linear multielectrode array (figure 1(a)). When a stimulus train was delivered through the electrode, a highly stereotyped stimulation artifact appeared on all channels of the recording array (figure 1(b)). Amplitudes could exceed ±7500 \(\mu\)V which is nearly two orders of magnitude larger than spiking waveform amplitudes (≈100 \(\mu\)V). The artifact traces resembled filtered versions of the original current stimulus; similar transients were evoked for each of the 20 biphasic pulses. These artifacts increased in amplitude with larger stimulation current and smoothly varied over the 24 channels of the recording probes (figures 5(a) and (b)).

4.1. Array location inference from artifact size

We used these amplitude profiles over channels for each session to fit a model relating artifact amplitude and the reciprocal of the distance of each electrode to the stimulation source. The model jointly optimizes a scaling parameter of the
amplitude-distance relationship as well as parameters specifying the location of the recording probe on each session (See Methods). We fit this model to data collected in nine recording and stimulation sessions in macaque prefrontal cortex. Figure 5(c) shows the recorded artifact amplitude (normalized by stimulation current) under the model against the model prediction, in which artifact size falls off with distance from the stimulation site (squared correlation $R^2 = 0.91$). The fitted probe locations are plotted in figure 5(d); these closely aligned with the approximate locations noted during probe insertion. We note that ERAASR itself does not assume any particular relationship among the artifact sizes among electrode channels; instead, ERAASR uses PC regression to identify and remove shared structure over the array channels. The close fit of the model to the empirical artifact size is consistent with a distance-dependent attenuation of the electrical stimulus, but we expect that ERAASR could be applied to datasets with arbitrary shared artifact structure among channels.
4.2. Artifact removal via ERAASR

We applied ERAASR to remove the artifacts from these traces. The ERAASR algorithm leverages several features of these recording datasets. First, it assumes that the true neural signal is corrupted by an additive stimulation artifact, which depends critically on the assumption that no information is lost due to amplifier saturation during the stimulation period. For the 5 μA–40 μA range of stimulation amplitudes used, the stimulation artifact produced did not saturate the amplifiers’ range (±8196 μV) on any of the recording channels in any experimental session. Therefore, if the artifact shape can be properly estimated, it can be removed via simple subtraction. Second, because we recorded simultaneously on multiple, closely spaced channels, we have multiple simultaneous observations of the electrical stimulation artifact at different points in space, whereas the individual channels do not share the same spiking waveforms, especially if we exclude immediately adjacent channels. Third, the stimulation train contains 20 repeats of identical biphasic pulses, which produced highly similar artifact transients. Finally, we deliver the stimulation train over many trials, yielding similar artifacts on each repetition. Because spiking activity evoked during the stimulation period is unlikely to occur at precisely the same time relative to each pulse and on each trial, we can use these repeated artifact measurements to build an estimate of the artifact shape for subsequent removal.

Figure 6(a) shows a set of cleaned traces for a single trial across the array on one of the recording sessions before high-pass filtering is performed; figure 6(b) shows the same data after high-pass filtering before spike detection. The raw, pre-cleaned voltage traces for this same trial are shown in figure 1(b), demonstrating that the vast majority of the artifact initially present has been removed. These high-pass filtered traces were then thresholded to extract spikes and then manually sorted into different units.

To evaluate the artifact removal process, we use prior knowledge about the multielectrode array channels. The most superficial (low numbered) and deepest (high numbered) channels were located above cortex and in white matter respectively, which we inferred from the signals observed on these channels, the density of neural signals on intermediate channels presumed to lie within cortex, and the known thickness of premotor cortex relative to the depth span of the probe. During non-stimulated trials, no detectable spiking waveforms were observed on these superficial and deep channels, therefore the presence of threshold crossings in the cleaned stimulation voltage traces on these channels would be surprising, and suggest that the cleaning procedure had failed to fully remove the stimulation artifact. However, as expected, the presence of evoked spiking activity during the stimulation window is limited to the intermediate channels, where spontaneous spiking activity was also observed (figure 6). For closer inspection, we superimposed the raw voltage traces with cleaned, artifact-removed, high-pass filtered voltage traces for individual trials. Figures 7(a) and (b) show such a comparison for the most superficial channel, where neither spontaneous nor evoked spiking activity was detectable. Figures 7(c) and (d) show the same comparison for a channel with spontaneous spiking as well as robust evoked spiking activity.

Next, it is essential to ensure that threshold crossings detected during the stimulation period are likely due to real neural spiking activity, as opposed to residual transients due to stimulation artifact that persists through the cleaning procedure. We reasoned that threshold crossings created by residual artifact on a specific channel would not be expected to resemble the spontaneous spiking waveforms collected during non-stimulated trials. For each session, we generated a visual comparison of the spiking waveforms detected on non-stimulated trials with those detected within the stimulation window (putatively ‘evoked’ spikes). These evoked spiking waveforms detected during the stimulation period were highly similar to the spiking waveforms detected during non-stimulation trials. Figure 8 demonstrates this similarity for a representative pair of experimental sessions. We note that these waveforms need not be identical, as evoked spiking activity...
Figure 7. Comparison of raw and artifact cleaned signals. (a) Raw (red) and cleaned (black) voltage traces for a single trial on the most superficially located recording channel where no spontaneous activity was recorded, and no spiking activity is detected during stimulation. (b) Zoomed view of (a). (c) Same as (a) but for a channel located within cortex where spontaneous activity was present. (d) Zoomed view of (c) demonstrating evoked spikes visible during the stimulation period.

Figure 8. Recovered spiking waveforms detected during the stimulation period closely resemble spontaneously occurring waveforms recorded on the same channel. ((a) and (b)) Superimposed waveforms from each channel with spiking activity of two representative recording sessions. Each pair of columns compares spontaneously occurring (spont.) versus peri-stimulus period (stim.) spiking waveforms. Waveform colors denote hand-sorted unit identity.
from other neurons located further from the electrode could superimpose to corrupt the waveforms from nearby neurons, especially when this faraway spiking activity is highly synchronized by stimulation.

We can also quantify the amount of residual artifact directly, by again utilizing channels that displayed no spontaneous spiking activity. Figure 9(a) compares the RMS voltage of these individual channels sampled from the post-cleaning stimulation window against a time window taken before stimulation. This metric captures the amount of baseline noise (largely thermal noise) on each channel, thereby providing an expectation for the variance of a given channel if the entirety of the corrupting artifact were successfully removed. For low amplitudes, the RMS voltage is often slightly lower than the pre-stimulation RMS, which is possible due to the greedy nature of the cleaning procedure. For larger amplitudes, the stimulation RMS remains acceptably close to pre-stimulation RMS. Figure 9(b) summarizes the similarity of spiking waveforms on each channel observed both spontaneously and during the stimulation window.

4.3. Effects of spike reliability and channel count

Fundamentally, ERAASR estimates electrical artifacts by exploiting the stereotypy of the artifact relative to variable spiking activity over channels, pulses, and trials. Over channels, this assumption is satisfied provided that the electrodes are spatially distributed enough to sample different populations of neurons; the regression step is performed so as to exclude channels sharing common spiking activity. Over pulses and trials, however, ERAASR relies on the assumption that evoked spiking activity is not perfectly reliable, such that spikes may be present on a subset of pulses with non-zero jitter in the pulse to spike latency. In order to better assess the distortion introduced for stably evoked activity that violates these assumptions, we performed a set of simulations. Here, we applied ERAASR to real datasets and extracted the estimated artifact on a single channel with no spiking activity. We then superimposed a known representative spiking waveform on this artifact at certain times, thereby generating synthetic datasets in which we varied the fraction of pulses with evoked spikes (the reliability) and the standard deviation of the spike timing relative to the previous pulse (the jitter). As expected, when spikes were inserted on nearly every pulse with a very narrow latency distribution, ERAASR recovered partially distorted versions of the true waveform (figures 10(a) and (b)) and failed to recover an increasing proportion of the spike times (figure 10(c)).

We also explored the dependency of channel count on ERAASR cleaning performance. Beginning with a real dataset, we applied ERAASR on subsets of channels with varying size. We then evaluated ERAASR performance by comparing the RMS of cleaned channels with no spiking activity, where the cleaned signal should be quiescent, against the RMS on those channels during non-stimulated trials. Figure 10(d) demonstrates that for these datasets, cleaning performance is poor and highly variable for low channel counts, but begins to saturate above 15 channels. However, additional channels would likely be more helpful in datasets where the channels exhibit more heterogeneity. Additional channels may also be beneficial when variability in the artifact over pulses and trials is increased, reducing the utility of these subsequent cleaning steps.

5. Discussion

We discuss the main features of our method for estimating and removing electrical stimulation artifacts in comparison with existing approaches, as well as limitations of our approach and possible improvements. We also highlight a set of interesting neuroscientific arenas where the ability to observe electrically perturbed neural activity might be particularly illuminating, underscoring the utility of artifact removal methods.

Here, we have applied the ERAASR algorithm to artifacts recorded on a linear multielectrode array (1D). However, we reiterate that ERAASR makes no assumptions regarding the electrode array geometry. The PCA regression step allows for artifacts on each electrode to be described as linear combinations of individual, shared components. Consequently, we anticipate that ERAASR should be readily applicable to 2D and 3D arrays with arbitrary geometry, provided that the artifact remains within the amplifier linear regime on the channels of interest such that no information is lost.

We also note that variability in electrode impedance should also not degrade ERAASR performance. The algorithm does not assume any relation between the artifact amplitudes over channels. Changes in the artifact size due to drifting impedance over experimental sessions could simply be dealt with by applying ERAASR independently on each dataset. If the artifact
5.1. Comparison to other methods

ERAASR exploits similarities in the electrical artifact observed across multiple channels, pulses within a stimulus train, and trials with repeated delivery of the same stimulus. Of these, the shared structure of simultaneously recorded artifact across multiple channels was most useful. Most previously described attempts exploit only the shared structure across trials, typically using an averaging or moving average approach (Hashimoto et al. 2002, Montgomery et al. 2005, Klink et al. 2017) to estimate artifacts on an individual channel. However, these approaches are necessarily quite sensitive to variability in artifact shape or amplitude across trials. While this variability may be small in relative terms, the much larger amplitude of artifacts relative to neural spiking waveforms can create meaningfully large errors in the residual signal, requiring that several milliseconds of the signal in the vicinity of the stimulation pulse be discarded. Another common approach is to exploit differences in the shape of stimulation artifacts relative to spiking waveforms, by using curve-fitting algorithms to capture artifacts in the voltage signal but exclude spiking activity (Wagenaar and Potter 2002, Erez et al. 2010). Similarly, these fits reliably capture artifact shape at a fixed delay from the stimulation pulse, but struggle to capture earlier portions of the transient (Erez et al. 2010), requiring peri-stimulus signal to be discarded. Motivated by the popularity of high-frequency pulse trains in ICMS experiments, our method addresses the problem of observing neural activity through the entire stimulation pulse without discarding signal.

Our algorithm combines a series of simple, intuitive, linear operations along the channels, pulses, and trials axes of the data tensor. The core operation is principal components regression, in which PCA is used to identify a template describing shared structure across a given axis of the tensor, followed by a regression step in which the artifact on a given channel (or pulse, or trial) is reconstructed by excluding that channel (pulse, trial) and its neighbors. This approach mitigates the risk of removing spiking signals during the cleaning process, exploiting the local nature of spiking responses across channels and the inherent biological variability in spiking responses on each pulse and each trial. Furthermore, each of these stages is intuitive and the results easily understood. A small set of key parameters, including the number of principal components and the set of neighbors excluded from the reconstruction at each stage, can be adjusted to tradeoff between more aggressive cleaning and more veridical preservation of spiking waveform shape.

Our method operationally separates the process of artifact estimation from spike detection and extraction, which is performed as usual using typical high-pass filtering and thresholding on all voltage data after the cleaning process is complete. In contrast, a promising alternative method for multi-channel artifact removal is presented by Mena and colleagues (Mena et al. 2016) to estimate artifacts on an individual channel.

Figure 10. (a)–(c) Recovery of spiking waveforms from synthetic datasets in which one channel was replaced by a generated signal with known spike waveforms and times. (a) Correlation of mean recovered waveform after application of ERAASR with the true waveform added to the signal, as a function of the fraction of pulses containing a spike and the jitter in the pulse to spike latency. Recovered waveforms displayed distortion as the spikes are evoked very reliably (higher fraction of pulses) and with very tight latency (lower jitter), as they become less distinguishable from artifact. (b) Mean waveforms recovered as a function of increasing jitter when all pulses contained a spike. (c) Fraction of spiking waveforms successfully detected as threshold crossings. (d) Quality of artifact removal when ERAASR was applied to subsets of channels of varying cardinality on the multielectrode array. RMS voltages are calculated only for channels which exhibited no spontaneous or evoked spiking activity and are normalized to each channel’s RMS voltage during non-stimulated trials. Error bars denote standard deviation over channels.
approach is highly effective and appropriate in the context of retinal recordings where individual neurons are recorded on many densely spaced electrodes, the electrical images of neurons in primate cortex using typical multielectrode arrays are often limited to one or two electrodes.

5.2. Limitations

First, our approach relies critically on the assumption that stimulation artifact linearly superimposes with neural spiking signals. This requirement is difficult to avoid; when the amplifier saturates, information about the spiking activity that would superimpose with the artifact is lost. Therefore, ERAASR’s utility is limited to stimulation configurations and amplitudes which occupy the linear, non-saturated regime of the amplifier, or at least to timepoints within the stimulation period where the artifact is not saturating. Practically, this sets a minimum distance between the electrode array and the stimulation source for a given stimulation amplitude. This distance could be designed such that the return path of the electrical current steers the electric field so as to minimize the recorded artifact amplitude (Rattay and Resatz 2004). This problem can also be effectively managed in hardware, employing special circuitry to estimate and partially cancel artifacts online (Brown et al. 2008, Wichmann and Devergnas 2011, Müller et al. 2013). A hybrid approach employing multi-channel artifact removal circuitry to prevent saturation using a predictable transformation of the recorded signal, supplemented with post hoc artifact cleaning procedure like the one described here could be effective across a much larger range of stimulation amplitudes.

Second, our algorithm sequentially removes shared structure across channels, pulses, and trials. We accomplish this by reshaping the voltage data tensor into a matrix (or set of matrices) so that familiar methods like PCA can be employed. This problem of finding shared structure naturally lends itself to tensor decomposition (Kolda and Bader 2009), which could identify shared structure jointly over each of these axes and potentially improve the artifact estimation. Tensors naturally arise in neuroscientific data collection contexts, and decomposition methods are becoming increasingly useful for identifying population structure (Seely et al. 2016, Elsayed and Cunningham 2017, Williams et al. 2017).

Third, our algorithm operates by greedily removing any shared structure as artifact, which can inadvertently remove or distort spiking waveform signals as well. In our dataset, we did not observe this while removing shared structure across channels, as we explicitly excluded adjacent channels from the reconstruction process. However, removing shared structure across pulses and trials assumes that spiking responses evoked by a stimulation pulse are variable in time. Moreover, with increasing stimulation amplitude, the temporal precision of evoked spikes may increase (Ranck 1975, Butovas and Schwarz 2003), which would lead to waveform distortion due to overzealous reconstruction and removal of spiking signals as artifact. We experimented with using matched filters for robust spike detection while relaxing the artifact estimation technique (data not shown), though we did not explore whether this would be applicable at higher currents and more synchronous evoked spiking. In these circumstances, we expect that a joint estimation of artifact and spiking as proposed by Mena et al. (2016) could be adapted to more effectively recover highly regular evoked spiking by exploiting other differences in the structure of spikes from artifact.

Fourth, the computational complexity of our algorithm could likely be improved. For a dataset with 24 channels, 60 ms time windows at 30 kHz, and 150 trials, the full alignment and cleaning procedure can be performed in approximately 75 seconds on a laptop. The computational time complexity of the ERAASR algorithm as implemented is O(C^3 + R^3 + P^3 + CRC^2 + PRTC^2 + RTP^2 + TCP^2) where C is channel count, R is trial count, P is pulse count, and T is timepoints per pulse. The algorithm comprises three rounds of PCA and linear regression on the rearranged artifact tensor. Faster, approximate versions of these algorithms exist and could be used to improve the computation time. To support online neural signal estimation, many of these steps could be precomputed and artifact subtraction applied to new observations one pulse at a time.

5.3. Motivation for direct observation of electrically perturbed neural activity

The initial conception of ICMS suggests that stimulation activates most neurons within a sphere surrounding the electrode tip (Stoney et al. 1968, Tehovnik 1996, Tehovnik et al. 2006). This idea derives originally from the findings of Stoney et al. (1968), who cleverly side-stepped the issue of stimulation artifacts by using collisions of antidromic and orthodromic spikes in a dual-stimulation paradigm as an indirect measure of local neuronal activation. A large body of subsequent work (e.g. Ranck (1975), Asanuma et al. (1976), Marcus et al. (1979), Tehovnik (1996), Rattay (1999), Nowak and Bullier (1996), Nowak and Bullier (1998a, 1998b), Kimmel and Moore (2007), Tehovnik and Slocum (2007)) has demonstrated that an individual current pulse primarily evokes spikes at axons—in particular, the excitable axon initial segment—and that the likelihood of evoking a spike in a neuron depends on distance, pulse duration, and current according to a relatively simple power law. The spatial localization of ICMS is consistent multiple behavioral studies as well (e.g. Murasugi et al. (1993), Tehovnik and Sommer (1997), Tehovnik et al. (2004, 2005), Bartlett et al. (2005)). These studies derive an estimate of the effective activation volume and dependency on stimulation parameters by combining behavioral readouts with known features of cortical spatial organization. For example, Murasugi et al. (1993) found that low current ICMS delivered to monkey area MT could bias perceived dot motion direction, but that ICMS currents above 20 μA less effectively biased perception. The authors conclude that below this threshold current, electrical activation is confined to a single ‘directional’ column in MT, about 0.2 mm in diameter (Albright et al. 1984).

However, recent evidence challenges this traditional view of localized, dense activation. Using two-photon calcium imaging, Histed et al. (2009) demonstrated that low-intensity ICMS activated a sparse collection of neurons distributed in
a relatively large volume extending several hundred μm in mice, and as far as 4 mm away in cats with 10 μA stimulation, perhaps resulting from lateral axons extending multiple millimeters within layer 2/3 of visual cortex. By advancing the electrode 30 mm, the authors found that an entirely non-overlapping sparse set of cells was activated, suggesting that ICMS at low currents (10–25 μA) activates axonal processes within ≈15 μm of the electrode tip. The directly activated population is therefore potentially sparse, widely distributed, and stochastic due to a precise dependence on axonal proximity to the stimulating electrode.

Many mapping studies have focused on direct activation in response to a single isolated current pulse, whereas multiple pulses at high frequency are much more common in scientific and prosthetic experiments. When longer pulse trains are used, particularly at high currents, electrical stimulation might readily engage a large collection of neural circuits while driving temporally complex neural responses (Graziano et al 2002, Strick 2002). ICMS induces transsynaptic spiking in connected neurons (Stoney et al 1968, Butovas and Schwarz 2003), and as larger currents activate more neurons, the effect of transsynaptic activation may be enhanced due to the summation of many from this larger directly activated population. Cortically evoked spiking may also depend on subthreshold activity in neurons, which is continuously modulated by ongoing task or stimulus-evoked responses (Kara et al 2002). Short-term synaptic depression commonly seen in cortex would also modulate the effectiveness of transsynaptically evoked spiking (Deisz and Prince 1989, Thomson and Cook 2010). Additionally, in an experiment combining cortical stimulation with fMRI and extracellular recording, Logothetis et al (2010) conclude that high-frequency ICMS disrupts normal information flow along cortico-cortical pathways, by recruiting strong inhibition that prevents the spread of activation beyond the regions whose direct afferents were stimulated. Consequently, a population of cells perturbed electrically may undergo a complex, time-varying pattern of modulation that reflects circuit microarchitecture, local circuit dynamics, short-term synaptic plasticity, recruitment of inhibition, and an interaction with ongoing task-related activity.

With this evident complexity, directly recording the effect of ICMS on neural activity will likely clarify the circuit mechanisms underlying intriguing aspects of the interaction between microstimulation and behavior. An accurate picture of this interaction could address many longstanding questions, such as how electrically evoked eye movements interact with visual guided eye movements (Sparks and Mays 1983), how precise timing of electrical pulses modulates saccadic effects (Kimmel and Moore 2007), how perceptual decisions are biased by stimulation of LIP (Hanks et al 2006), how motor preparation recovers subsequent to electrical disruption (Churchland and Shenoy 2007, Shenoy et al 2011), and how motor cortical stimulation can evoke complex movements converging on a specific endpoint (Graziano et al 2015) and effectively nullify the contribution of goal-directed behavior in the motor cortical output (Griffin et al 2011). A detailed comparison with the effect of optogenetic stimulation could clarify the differences (Diester et al 2011, Gerits and Vanduffel 2013, Ohayon et al 2013) and similarities (Dai et al 2013) observed between the two modalities’ effects on behavior. Lastly, a detailed model of microstimulation’s interaction with cortical dynamics could facilitate delivery of more reliable or realistic artificial sensory perception in visual, auditory, and motor prostheses (Otto et al 2005, Teohvnik et al 2009, O’Doherty et al 2011, Bensmaia and Miller 2014, Dadarlat et al 2015).

6. Conclusion

We developed ERAASR, an algorithm for estimating and removing artifacts caused by electrical stimulation on multi-electrode array experiments. ERAASR assumes that artifact is shared over many channels and that evoked transients are highly repeatable across pulses and trials, whereas spiking activity is highly local and temporally jittered. Shared structure in the voltage signals is identified and removed sequentially across channels, across pulses in a stimulus train, and across trials, using straightforward linear methods. We believe our technique will be useful to neuroscientists in drawing precise causal links between perturbations and the effect of stimulation on neural activity and behavior and aid in the design and implementation of enhanced neural prosthetic systems capable of restoring lost sensation and facilitating precise control.

7. Code availability

Full Matlab code for ERAASR is available at https://github.com/djoshea/eraasr along with a link to a representative example dataset.

Acknowledgments

We thank Tirin Moore, Rob Franklin, and Paul Venable for helpful discussions. We thank Peiley Lau for devising the ERAASR acronym. This article was supported by the following grants: Christopher and Dana Reeve Paralysis Foundation, Burroughs Welcome Fund Career Awards in the Biomedical Sciences, Defense Advanced Research Projects Agency Reorganization and Plasticity to Accelerate Injury Recovery N66001-10-C-2010, US National Institutes of Health Institute of Neurological Disorders and Stroke Transformative Research Award R01NS076460, US National Institutes of Health Director’s Pioneer Award 8DP1HD075623-04, US National Institutes of Health Director’s Transformative Research Award (TR01) from the NIMH 5R01MH09964703, Defense
Advanced Research Projects Agency NeuroFAST award from BTO W911NF-14-2-0013, Howard Hughes Medical Institute, NSF GRFP, and NSF IGERT.

ORCID iDs

Daniel J O’Shea  https://orcid.org/0000-0002-1366-1743
Krishna V Shenoy  https://orcid.org/0000-0003-1534-9240

References

Asanuma H, Arnold A and Zarzecki P 1976 Further study on the excitation of pyramidal tract cells by intracortical microstimulation Exp. Brain Res. 26 443–61
Chung S and Ferster D 1998 Strength and orientation tuning of the thalamic input to simple cells revealed by electrically evoked cortical suppression Neuron 20 1177–89
Deisz R A and Prince D A 1989 Frequency-dependent depression of inhibition in guinea-pig neocortex in vitro by GABAB receptor feedback on GABA release J. Physiol. 412 513–41