Published in final edited form as:

J Neural Eng. 2018 April 01; 15(2): 026014–. doi:10.1088/1741-2552/aa9ee8.

Signal processing methods for reducing artifacts in microelectrode brain recordings caused by functional electrical stimulation

D. YOUNG^{*,1,2}, F. WILLETT^{1,2,3}, W. D. MEMBERG^{1,2}, B. MURPHY¹, B. WALTER^{4,5}, J. SWEET^{6,7}, J. MILLER^{6,7}, L. R. HOCHBERG^{8,9,10,11,12}, R. F. KIRSCH^{1,2}, and A. B. AJIBOYE^{1,2}

¹Case Western Reserve Univ., Cleveland, OH

²FES Ctr. of Excellence, Rehab. R&D Service, Louis Stokes Cleveland Dept. of VA Med. Ctr., Cleveland, OH

³Dept of Neurosurgery, Stanford University, Stanford, CA

⁴Neurol., UH Cleveland Med. Ctr., Cleveland, OH

⁵Neurol., CWRU Sch. of Med., Cleveland, OH

⁶Neurosurg., UH Cleveland Med. Ctr., Cleveland, OH

⁷Neurolog. Surgery, CWRU Sch. of Med., Cleveland, OH

⁸Ctr. for Neurorestoration and Neurotechnology, Rehab. R&D Service, Dept. of VA Medical Center, Providence RI

⁹Sch. of Engin., Brown Univ., Providence, RI

¹⁰Neurol., Massachusetts Gen. Hosp., Boston, MA

¹¹Neurol., Harvard Med. Sch., Boston, MA

¹²Inst. For Brain Sci., Brown Univ., Providence, RI

Abstract

Objective—Functional electrical stimulation (FES) is a promising technology for restoring movement to paralyzed limbs. Intracortical brain-computer interfaces (iBCIs) have enabled intuitive control over virtual and robotic movements, and more recently over upper extremity FES neuroprostheses. However, electrical stimulation of muscles creates artifacts in intracortical microelectrode recordings that could degrade iBCI performance. Here, we investigate methods for reducing the cortically recorded artifacts that result from peripheral electrical stimulation.

*Corresponding author (dry5@case.edu).

Author Contributions

DY contributed in study conception, designed the experiments, performed data analyses, and drafted the manuscript, which was further edited by all authors. FRW assisted in design of experiments, reduction techniques, and analysis methods. WDM assisted in conception and design of FES protocols and performed data collection. BM performed data collection with participant T8. BW provided the participant with post-surgical care. JS and JM planned and executed the neurosurgical placement of the electrode arrays. LRH was the investigational device exemption sponsor-investigator of the pilot clinical trial. ABA and RFK assisted in study conception and design, and interpretation of results.

Approach—One participant in the BrainGate2 pilot clinical trial had two intracortical microelectrode arrays placed in the motor cortex, and thirty-six stimulating intramuscular electrodes placed in the muscles of the contralateral limb. We characterized intracortically recorded electrical artifacts during both intramuscular and surface stimulation. We compared the performance of three artifact reduction methods: blanking, common average reference (CAR) and linear regression reference (LRR), which creates channel-specific reference signals, composed of weighted sums of other channels.

Main Results—Electrical artifacts resulting from surface stimulation were 175x larger than baseline neural recordings (which were 110 μ V peak-to-peak), while intramuscular stimulation artifacts were only 4x larger. The artifact waveforms were highly consistent across electrodes within each array. Application of LRR reduced artifact magnitudes to less than 10 μ V and largely preserved the original neural feature values used for decoding. Unmitigated stimulation artifacts decreased iBCI decoding performance, but performance was almost completely recovered using LRR, which outperformed CAR and blanking and extracted useful neural information during stimulation artifact periods.

1. Introduction

The present study investigates several methods for reducing stimulation artifacts in intracortical recordings during functional electrical stimulation (FES) of muscles in the arm and shoulder, with the goal of enhancing decoding of movement commands from intracortical brain-computer-interfaces (iBCIs). FES can be used to restore functionally meaningful movements, such as reaching and grasping [1,2], to persons with extensive paralysis resulting from spinal cord injury (SCI). In the absence of voluntary motion, iBCIs offer a useful command interface for persons with chronic tetraplegia to control artificially actuated functional activities, such as control over computer interfaces [3–5], reaching movements using a robotic arm [6–8] or continuous grasping using their own hand actuated through a surface arm sleeve [9]. Recently, our group showed for the first time a person with tetraplegia performing functionally meaningful iBCI-commanded reaching and continuous grasping movements, using their own arm and hand, actuated through an intramuscular FES system [10].

Previous studies have demonstrated that electrical stimulation can create artifacts (stimulation-induced waveforms that mask underlying signals of interest) on even distant recording electrodes in a variety of applications, including deep brain stimulation [11], electroencephalograms [12], and electromyography [13]. While iBCIs rely on precise recordings of microvolt-scale signals, FES generates relatively larger electric fields in the paralyzed limbs that can propagate through the body to the cortex. iBCI recordings during FES can thus be contaminated by relatively high amplitude electrical artifacts that distort underlying neural information and typically bias neural features extracted by signal processing towards higher values [9]. Despite their magnitude and potentially strong negative impact on performance, these artifacts remain largely uncharacterized. Specifically, it is unclear 1) how these artifacts differ between stimulation techniques (surface [9] or intramuscular [10] FES), 2) how they affect typical neural features and decoding

performance, and 3) what methods are appropriate for reducing artifact corruption during iBCI control.

The majority of artifact removal methods implemented in biomedical applications can be grouped into three main categories: specialized hardware, blanking, and signal processing techniques.

Specialized hardware is usually developed for applications with large artifacts that saturate amplifiers or when latency in recording is unacceptable, such as for closed-loop deep-brain-stimulation. Two reported solutions have focused on applying an analog filter to the signal prior to final amplification [14,15] or on using differential recording electrodes and also saturating voltages above a set threshold [16]. However, hardware artifact reduction is usually application specific, expensive to develop, and can also irreversibly alter the desired neural activity (through filtering or saturation).

Blanking, the exclusion of data during stimulation and artifact periods, is the most common artifact removal technique. It can be implemented in hardware using periodic recording [17] or sample-and-hold techniques triggered only during stimulation [18]. Blanking can also be implemented in software, where time periods with stimulation artifacts are either removed [9,19], or replaced with constant [13,20,21], interpolated [22] or simulated data [23]. However, blanking is often not be an optimal artifact reduction technique in iBCI applications, where the removal of data ignores important information for substantial periods following stimulation and thus reduces performance, a particularly relevant issue when many different muscles are being stimulated.

Alternatively, signal processing based artifact reduction techniques attempt to retain data during stimulation by reducing the magnitude of artifacts with a variety of software methods. When the artifacts reside in a different frequency range than the desired signal, simple software filtering can be used [24]. In other applications, artifact properties do not change over time, so template subtraction techniques which detect stimulation periods and remove an average artifact waveform [11,21,25] can be used. If the artifact can be predicted based on the stimulation parameters, methods such as curve fitting can be used to build models of the artifact waveforms [26]. Traditional spike sorting techniques can also be used in some applications to separate artifact from signal [27]. However, spike sorting as a means to reduce stimulation artifacts will likely not recover any activity during stimulation periods, resulting in performance similar to blanking. Additionally, this method will only work when spike sorting is explicitly part of the signal processing methodology, and most iBCI applications now use multi-unit threshold crossings or continuous high frequency power measurements as features [9,10,28,29].

When considering methods for artifact removal from neural recordings for an FES+iBCI system, a method that records and maintains as much neural signal as possible is preferred, as it should provide the most accurate estimate of the users intended movements. Therefore, artifact reduction techniques that recover information during stimulation periods are expected to result in higher performance than blanking alone. Additionally, there are some unique properties of FES+iBCI systems to consider when choosing an artifact reduction

technique. First, FES stimulation parameters (e.g. pulse width) can vary in time, with each electrode having a customized stimulation pattern which varies widely depending on the arm and hand configuration. This makes traditional template subtraction or static modeling techniques very difficult and onerous to implement because of the enormous number of possible stimulation combinations and thus the time-consuming effort needed to characterize a huge library of templates. Additionally, iBCI recordings use high-density microelectrode arrays (96 channels, $4 \times 4 \text{ mm}^2$ area) to sample from electrodes that likely share nearly identical artifact responses, given the distant stimulation source. This led us to explore two artifact reduction techniques that utilize information from other channels within an array to identify and remove the artifact waveforms.

In this study, we characterized the stimulation artifacts resulting from both intramuscular and surface FES. We implemented three software based artifact reduction techniques: blanking, common average reference (CAR), and linear regression reference (LRR). The techniques were compared using ‘offline’ analysis, where each method was applied to the same pre-recorded blocks, to eliminate block-to-block variability in neural signals, patient fatigue, etc. The same techniques can be readily implemented for online use as well. By applying each technique to the same data, we directly compared 1) artifact magnitudes, 2) changes in neural features, and 3) the effect of artifacts on decoding performance. Results demonstrated that FES artifacts increased extracted neural features values and reduced decoding performance during an iBCI task. We show that the LRR method outperformed blanking and CAR, and achieved >90% normal decoding performance during surface stimulation periods and nearly full decoding performance during intramuscular stimulation periods.

2. Methods

Permission for these studies was granted by the US Food and Drug Administration (Investigational Device Exemptions #G090003 and #G950116) and the Institutional Review Board of University Hospitals Cleveland Medical Center (protocol #04-12-17) and Massachusetts General Hospital (2011P001036). The participant was enrolled in a pilot clinical trial of the BrainGate Neural Interface System ([ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT00912041), number NCT00912041). Informed consent, including consent to publish, was obtained from the participant prior to his enrollment in the study.

Participant and Implants

We studied one participant (ID = T8) who was enrolled in our study 8 years after a traumatic spinal cord injury (C4 level, AIS A) resulting in chronic tetraplegia. T8 received two 96-channel intracortical microelectrode arrays (Blackrock Microsystems, Salt Lake City, UT, USA) placed in the hand/arm area of the left motor cortex. In two follow-up procedures (trial days 125 and 280), T8 received 32 percutaneous intramuscular stimulating electrodes and 4 intramuscular anodes (316LVM stainless steel insulated with fluorinated ethylene propylene, Synapse Biomedical, Oberlin, OH, USA) placed in the upper and lower right arm and shoulder (figure 1). The stimulating electrodes (10 mm^2 approximate stimulating surface area) targeted muscles for restoring finger, thumb, wrist, elbow, and shoulder movements,

while the anodes (100mm² approximate surface area) were placed in soft tissue in the vicinity of the stimulating electrodes for each of the four percutaneous connectors [10].

Neural Recordings

Each intracortical array was attached to a percutaneous pedestal connector on the head of the participant. Patient cables connected the pedestals to amplifiers (Blackrock Microsystems), which filtered (0.3Hz – 7.5kHz bandpass) and digitized (30kHz) the neural signals. The amplifiers have an input voltage operating range of $\pm 8.095\text{mV}$, an input impedance $>10^{12}$ ohms, and a resolution of 16 bits at 250nV / bit. An output trigger signal from our stimulation device indicated periods of stimulation and was recorded by the Neuroport Signal Processor (Blackrock Microsystems) for synchronizing with the neural data.

Stimulation Parameters

At the beginning of each session, temporary surface patch electrodes (surface area 6.25cm²; Axelguard Manufacturing Co, Fallbrook, CA, USA) were placed to activate similar muscle groups to those implanted with the intramuscular electrodes (Figure 1(B)). Stimulation was delivered via a custom stimulator (Universal External Control Unit (UECU); Cleveland FES Center, Cleveland, OH, USA), which was battery powered and isolated from our recording equipment. The UECU could stimulate up to 24 intramuscular channels, which were attached via cables to four multi-channel percutaneous connectors that also included one implanted anode per connector. Only four total channels of bipolar surface stimulation could be connected to the UECU at one time. Intramuscular stimulation parameters were fixed at 20mA, 12.5Hz (80ms inter-stimulation period), and varied between 0-250 μs pulse width. Surface stimulation required increased current due to increased skin impedance, so the amplitude was fixed at 60mA. Stimulation waveforms were charge-balanced biphasic pulses consisting of a cathodic square wave, a short delay, and a lower amplitude anodic charge recovery phase. The UECU was programmed to stimulate a maximum of two channels simultaneously, with a 1ms delay between consecutive pulses to different channels, and all pulses being delivered within 5-8ms. Desired pulse widths for every channel were sent to the UECU from a computer running xPC Target, Matlab's real time operating system (The Mathworks Inc, Natick, MA, USA), and updated every 100ms.

Experimental Design

Our primary goal was to compare the ability of artifact reduction techniques to reduce artifact size, recover useful information from ongoing neural activity during stimulation periods, and enable accurately decoded iBCI commands. To perform decoder accuracy analyses, we collected a dataset containing FES artifact contamination during completion of a typical iBCI virtual task (not iBCI-commanded FES movements), which allowed us to measure decoding performance during stimulation and non-stimulation periods. Artifact reduction technique comparisons were performed 'offline' (each method was applied post hoc and performance measured on the same data) to eliminate block-to-block variability in neural signals, participant fatigue, etc. We deliberately did not use data of iBCI-commanded FES movements, because those artifacts could contain task-related information (which contacts are activated and at what strengths conveys information about the current limb

position, likely future movements, etc.) which could spuriously improve offline decoder performance.

Three sessions of iBCI tasks with randomized and task-unrelated functional electrical stimulation were collected (trial days 582, 585, 623). In each session, the participant performed a 3D virtual reality iBCI task where he was instructed to attempt movements to command a virtual arm towards a target position displayed on screen (Supplementary Figure 1). While the participant performed the virtual reality task, his arm and hand were stimulated using either intramuscular or surface patch electrodes. The participant was instructed to ignore his limb movements and concentrate on the virtual task, and, due to his paralysis, he reported no sensation of the FES generated movements. The stimulation commands consisted of replayed pulse-width trains from previous elbow, wrist, and hand control FES tasks [10] and produced movements that were uncorrelated with movements of the virtual arm. The reconstructed movements in each block, which included movement of a motorized mobile arm support, spanned a wide range of joint and limb positions that are typical to reaching and grasp movements. Between 8 and 17 electrodes were stimulated in a block using intramuscular electrodes (depending on the replayed task); due to hardware constraints, only four electrodes were stimulated during surface blocks. The participant was well trained in the virtual reality task, achieving >90% success rate across several sessions and months. During a session, the participant repeated the iBCI virtual reaching task every block while we alternated between three FES conditions: no-stimulation, intramuscular stimulation, and surface stimulation.

The session began with our standard virtual-control training paradigm, beginning with two blocks of open-loop attempted movements, where the computer displayed predefined movements that the participant attempted to make. Next, the participant performed two blocks of closed-loop control, where the virtual limb movements were directly controlled by neural activity, with movement error attenuation, i.e. computer assistance that reduced off-target movements [30]. After completing these four training blocks, a final decoder to map the neural activity to virtual arm movements was built and used without movement error attenuation for the remainder of the session. For closed-loop iBCI control during FES stimulation, stimulation was applied at 12.5Hz (80ms inter-stimulation period) with the pulses to the various muscles occurring during the first 5-8ms of each 80ms period. Online blanking of the intracortical recordings was conservatively implemented for a 17ms period, initiated by a trigger signal from the UECU. This allowed the participant to complete the task with normal success rates (mean 93%), even with constant stimulation.

In one additional experimental session (trial day 562), single-electrode FES stimulation was performed without any iBCI virtual task. The goal of this session was to record data that characterized the stimulation artifact by repeating the same stimulation parameters on a single channel throughout a one-minute block. These data were used to more accurately report average artifact waveforms and amplitudes, since the parameters from replayed FES tasks utilized time varying pulse-widths across stimulating electrodes.

Artifact Reduction Methods

We chose to implement and compare three different artifact removal techniques: blanking, common average reference (CAR), and linear regression reference (LRR). Blanking and CAR are the prevailing noise reduction methods in iBCI applications, and LRR was chosen because it is a data-driven approach that can be easily trained on user and application-specific artifacts. Each chosen method can be easily implemented in software on the digitized neural recordings, requiring minimal computation time and enabling their use in low-latency closed-loop BCI applications. Importantly, all three methods are applied to the continuous voltage recordings, making them effective regardless of the chosen iBCI features. Table 1 describes the three techniques and their primary assumptions about the artifact noise.

Blanking makes the most aggressive assumption: that the data is so corrupted during stimulation periods that it should be discarded completely (Table 1). We blanked a continuous period of 5ms (surface) or 8ms (intramuscular), determined by calculating artifact durations from actual multi-electrode stimulation responses across multiple stimulation parameters and microelectrode channels. During the blanking period, the recorded data was replaced with linearly interpolated points (connecting the last point of pre-blanked data with the first point of post-blanked data), which reduced the smearing effect of bandpass filtering the large amplitude artifacts into nearby timepoints. However, when computing the neural features, all ‘blanked’ timepoints were excluded and features were calculated using only the non-artifact periods in each 20ms window.

Common average reference (CAR) attempts to recover underlying neural information during stimulation periods by re-referencing each electrode to remove common noise across all channels (Table 1). The new reference is usually built using an equally weighted average of all channels on the array, and then applied by subtracting the reference from each channels raw voltage recordings. In this study, only 80 of the 96 channels on each array were used to build the CAR, determined by lowest variance during a 1 minute baseline recording at the beginning of the session, consistent with previous applications [29]. Using a CAR for artifact removal assumes the noise waveforms are the same across channels and cannot remove any channel-specific artifacts.

The linear regression reference (LRR) method also attempts to remove noise artifacts by re-referencing channels to a combination of the other channels on the array. However, LRR does not assume the noise waveforms are equivalent across all channels, and instead builds a unique reference signal for each channel, enabling channel-specific variations in noise to be removed (Table 1). The weight coefficients for each channel are derived from a linear regression predicting a given channel’s activity using recordings from the other 95 channels as features. The method was proposed previously by Musial et al. as a technique for reducing typical microelectrode noise sources (line noise, EMG, field potentials, etc.) to improve spike sorting performance [31]. For our implementation, the channel references were trained using only data contaminated by stimulation artifacts, in an attempt to best reduce the largest source of noise in our system. The output of the regression was a set of weights that, when applied to the recordings of the other 95 channels on the microelectrode array, optimally recreated the stimulation artifact periods on that given channel. The weights

were then applied to the entire block of recordings, creating a continuous reference signal that was subtracted from the channel's recorded voltage.

For the LRR method, we modeled the recorded voltages on a given channel (R_c) as a linear sum of the voltages on the other channels (X) plus signal (S_c).

$$R_c = X \vec{w}_c + S_c$$

R_c are S_c and $t \times 1$ vectors and X is an $t \times (N - 1)$ matrix. t is the number of data points in the training dataset and N is the number of channels on a given array (96). We included every channel within each array and performed the regression separately for each of the two arrays. $\vec{w}_c$ is an $(N - 1) \times 1$ vector of weights that, when multiplied by X , best recreate R_c .

The weight vector $\vec{w}_c$ was estimated using least squares regression, $\vec{w}_c = (X^T X)^{-1} X^T R_c$.

When evaluating the LRR method on any particular block, we used weighting vectors trained using data from all other blocks from the session with the same FES stimulation type so that the weighting vectors were “cross-validated”. For training, we included all data points from each stimulation trigger pulse that were above a voltage threshold, determined manually for each stimulation type based on average artifact waveforms. This resulted in an average training data duration of 26.4s (downsampled 15k samples/sec * 26.4s = 396k data points) for intramuscular and 70.0s (15kps * 70s = 1.05M data points) for surface stimulation blocks.

Offline Feature Computation

Artifact reduction techniques were applied ‘offline’, where the voltages that each method would have produced within a session were recreated using pre-recorded data. To accomplish this, computed neural features were reconstructed for each artifact reduction method using the raw 30kHz recordings from each electrode and typical feature construction methods from recent literature [10,29,32,33]. Consistent with previous methods, a 15kHz low pass filter (30 order Hamming window) was applied to the recorded (and 0.3Hz – 7.5kHz bandpass analog filtered) signal, then downsampled to 15kHz. We then optionally applied one of our three artifact reduction methods (blanking, CAR, or LRR). Next, a non-causal Butterworth bandpass filter (8th order) between 250Hz to 5000Hz was applied. From this spike band signal, two neural features were extracted: threshold crossing rate (TX) and high-frequency spike-power (HFSP). Threshold crossing rate is defined as the number of threshold crossings counted in a window, divided by the duration of the window, with units of [# crossings / sec]. A threshold crossing was triggered by a negative deflection past a channel-specific voltage threshold, set at -4.0 times the RMS amplitude of the channel during a one-minute reference block collected at the beginning of the session. High-frequency spike power was calculated as the mean of the squared voltages of the spike band signal within a time window, with units of [μV^2]. The procedure resulted in two features per channel, and with 96 channels on each of our two arrays, producing $2 \times 96 \times 2 = 384$ total features.

During normal system performance, these neural features were calculated in 20ms non-overlapping windows. When blanking, the blanked periods (5-8ms) were excluded from later feature computation steps and the computations were adjusted accordingly for the reduced window size. Some offline analyses (labeled as “artifact period only”) used windows of shorter duration for isolating performance only during periods with known artifact contamination (called “artifact only” bins).

Decoding Methods

Decoders were constructed using data from all closed-loop non-stimulation blocks within each session. Unique decoders for each of the three artifact reduction methods were built, using feature vectors computed after the reduction method was applied to the non-stimulation data (Figure 2(A)). When applying LRR to non-stimulation data, the reference matrix was trained using an equivalent duration of continuous data from each channel and this activity was regressed against the recordings from the other 95 channels on the array. When building decoders for the blanking condition, blanking was not artificially applied to the training data; this allowed as much training data as possible to be included and resulted in the blanking decoders being equivalent to the Raw condition decoders. Decoders were trained using all possible features (TX and HFSP from each channel), which were z-scored prior to training.

To build decoders, neural features were regressed against a three dimensional point-at-target vector that pointed directly from the cursor position to the target, with a constant magnitude. A variant of optimal linear estimation (OLE) was used to produce a decoding matrix D that mapped the 384 neural features to intended velocities (full methods described in [10]). The decoder was applied at each timestep using the equation: $u_t = Df_t$, where the decoder matrix D (size 3×384) was applied to a vector f_t of neural features (size 384×1), yielding the commanded velocity u_t (size 3×1). For online control, this command vector was then smoothed with a first-order low pass filter [10]; for offline analysis, only the unsmoothed vector u_t was considered. A 400ms reaction time interval after target presentation was excluded from our training data.

We used a decoder signal-to-noise ratio (SNR) metric for evaluating the performance of the artifact reduction techniques, following the methods reported in (Willett, Murphy, et al. 2017, Supplemental Section 3). First, we modeled the decoded velocity as a sum of an ‘intended velocity’ and noise: $u_t = c_t + e_t$, where u_t is the decoded velocity, c_t is the assumed intended velocity, and e_t is the noise. We assumed the intended velocity was a position error vector, scaled by a non-linear function of target distance:

$$c_t = \frac{g_t - p_t}{\|g_t - p_t\|} f_{targ}(\|g_t - p_t\|),$$

where g_t is the target position, p_t is the cursor position, and f_{targ} is a nonlinear, scalar weighting function that was fit empirically. The inclusion of the scalar f_{targ} term allowed our decoder SNR measurements to account for smaller commanded velocities near the target, and was empirically fit on decoded velocities using four-fold cross validation. A full model

of user intention was first developed in [33], including additional terms for user feedback delay and cursor velocity. The simplified model shown above (including only f_{targ}) was later used for iBCI decoder SNR calculations in [32] and is appropriate whenever the smoothing dynamics are moderate and the cursor velocity term is not needed. Decoder SNR was defined as the magnitude of the signal divided by the standard deviation of the noise, where the signal is the intended velocity c_t and the noise is the decoder noise e_t (figure 2(B)). The magnitude of the signal was defined as the average magnitude when the user was far from the target (>80% away). We compared our artifact-reduced decoding performance to a ‘no-stim baseline’, which was chosen as the CAR decoder performance during non-artifact periods within the same block. We chose CAR as our baseline condition because it is standard in BMI feature computations, and because it was our highest performing method, with non-stimulation period results that were very similar to LRR. To test the significance of the decoder SNR measurements, each artifact reduction method was compared to the baseline condition using a paired-sample t-test with $\alpha = 0.05$ and the resulting p-values were adjusted for four comparisons using the Holm-Bonferroni method. The null hypothesis was that the mean of the sample differences was zero.

3. Results

Artifact Characterization

Muscle stimulation using either intramuscular or surface electrodes produced electrical artifacts on all intracortical recording channels. Figure 3(A) shows a typical channel response to single electrode stimulation during a one-minute block with 40s of stimulation (12.5Hz, 200 μ s). The stimulation period can be seen in the raw recordings as a large increase in magnitude compared to baseline. The properties of the stimulation artifacts were quantified by first calculating the stimulation-triggered average artifact waveform for each condition (figure 3(B)). During surface stimulation, the artifact waveforms were 3250 μ V_{pp} (peak-to-peak), >30 $\times$ larger than the baseline neural activity (110 μ V_{pp}, calculated as six times the standard deviation during non-stimulation periods). During intramuscular stimulation, the average artifact on this channel was 255 μ V_{pp}, >2 $\times$ larger than the baseline recorded magnitudes. Artifact durations were short (<1ms) for both stimulation conditions, and highly consistent. The trial-averaged artifact waveforms (*average baseline activity* < 1 μ V_{pp}) were highly correlated between channels within each array, with >95% of the channels having a greater than 95% median correlation to other channel waveforms (figure 3(C)).

LRR Weight Calculations

LRR weight matrices were built for each block by using linear regression to predict artifact period activity on each channel using the other channels. Figure 4 shows across-block average weight matrices (z- scored per channel and block to scale coefficients) for a few representative channels from the medial intracortical array. The weight matrices for intramuscular blocks tended to have very small contributions from most channels, with only a few nearby channels contributing most to the reference signal. Surface blocks typically had strong contributions from the same channels as the intramuscular condition, but also included moderate contributions from additional channels dispersed across the entire array

(figure 4; Ch. 136, 124, 170). A smaller number of channels had more unique patterns: Ch. 158's weight matrices were very similar across stimulation conditions, and Ch. 148's intramuscular weight matrix had moderate contributions from many more electrodes than usual. Average weight matrices for all channels are shown in supplementary figure 2. Broadly, there were strong contributions to the channel-specific references from a few of the closest channels, and the coefficient magnitudes dropped off substantially for more distant channels. This pattern held for both surface and intramuscular stimulation conditions, but was more variable for the further channels during surface stimulation blocks. Additionally, there are similar patterns evident within each array across the two stimulation conditions. This is congruent with expectations, as the strength of common noise between a pair of electrodes (and thus the relative magnitude of their LRR coefficients) is partially dependent on stable electrode and tissue properties and therefore should have some similarities across stimulation conditions.

The examples in figure 4 demonstrate the capabilities of the LRR method compared to the more commonly used CAR. The main advantage of the LRR is that, using actual noise properties from pre-recorded data, it solves for an optimal combination of channels to reject the stimulation artifact per channel, rather than utilizing a single template pattern across all channels. Despite the strong correlations in artifact waveforms across all channels (figure 3(C)), optimal prediction of the waveform on a given channel typically uses stronger contributions from the nearest channels rather than an even contribution across the entire array. However, initial results implementing a small Laplacian spatial filter [34], which references each channel to an average of its directly neighboring channels, showed performance similar to CAR and worse than LRR (supplementary figures 4-5). These results are consistent with the large variability in LRR weights across channels, and highlight the benefit of the LRR method which builds channel specific references using information from all channels on an array. In these experiments, LRR weight matrices were built for each block, even though matrices built on different blocks and days were quite similar. Averaged across all channels, LRR weights were highly correlated across blocks in the same session (mean $\pm$ standard deviation: $.88 \pm .09$ surface, $.97 \pm .01$ intramuscular), and between blocks from different sessions ($.85 \pm .08$ surface, $.97 \pm .01$ intramuscular).

Artifact Magnitude Reduction

The CAR, LRR, and blanking methods were applied offline and their effect on artifact magnitudes was measured. Each reduction technique was successful in reducing the magnitude of the stimulation artifacts (figure 5(A)). Figure 5(A) shows the average remaining artifact waveforms for each cleaning condition. Raw (uncleaned) recordings were characterized by large spikes during stimulation periods, with average artifacts $>3000\mu V_{pp}$ in magnitude. Applying the CAR reduced the size of the artifact spikes, with an average remaining artifact waveform $<300\mu V_{pp}$. LRR performed better than CAR at reducing the size of the artifact. After LRR cleaning, the large stimulation artifact spikes were no longer easily distinguishable from the filtered neural signal and the average remaining artifact was smaller in amplitude than background neural activity. After blanking was applied, the uncorrupted neural activity remained while data during stimulation periods was removed.

The average response to stimulation after blanking clearly showed the exclusion of information during the 5ms blanking period (figure 5(A)).

The cleaning performance of CAR and LRR compared to the raw recordings was summarized in both the time domain (quantifying artifact magnitudes) and frequency domain (using power spectral density plots). Blanking was omitted from these analyses because all data during artifact periods was completely excluded during blanking.

Artifact reduction performance was quantified in the time domain using average remaining artifact amplitude as a metric. Artifact magnitudes were calculated across all 192 channels using six blocks of surface and four blocks of intramuscular single-electrode stimulation data, which was better suited for artifact averaging because the stimulation parameters were held constant throughout the whole block (figure 5(B)). Without artifact reduction, both surface and intramuscular stimulation produced artifacts that were larger in magnitude (Surface= $3446\mu\text{V}_{\text{pp}}$ and IM= $225\mu\text{V}_{\text{pp}}$) than the baseline neural recordings ($110\mu\text{V}_{\text{pp}}$). Applying the CAR reduced the artifact size (Surface= $340\mu\text{V}_{\text{pp}}$ and IM= $36\mu\text{V}_{\text{pp}}$) for both stimulation conditions, but did not adequately clean the large surface stimulation artifacts. Remaining artifact magnitudes were smallest after LRR cleaning (Surface= $3.2\mu\text{V}_{\text{pp}}$ and IM= $2.3\mu\text{V}_{\text{pp}}$). The residual artifacts were $<10\mu\text{V}_{\text{pp}}$ for both stimulation types, indicating excellent performance at artifact removal using LRR on the single-electrode stimulation data.

We explored how the artifact reduction techniques affected the neural frequency content by computing the power spectral density (PSD) for every channel before and after artifact reduction (figure 5(C)).

During non-stimulation blocks, there was visible 60Hz harmonics noise, but both CAR and LRR greatly reduced that contamination. Surface stimulation increased signal power across a wide band of frequencies that overlapped with the spike band (250-5000Hz). Most contamination occurred at the stimulation frequency (12.5Hz) and its harmonics. CAR reduced some of the power introduced by the artifact but still left clear peaks corresponding to artifact contamination, while LRR performed better at restoring a PSD uncontaminated by stimulation artifacts. Intramuscular stimulation also increased power in higher frequencies compared to non-stimulation blocks. CAR reduced the high frequency power spikes, but still had identifiable noise in the relevant frequencies. LRR performed better at reducing the characteristic noise and produced a PSD very similar to that of non-stimulation data.

Neural Feature Reduction

In addition to characterizing reductions in artifact waveforms, we investigated the mechanism by which stimulation artifacts might affect iBCI performance by analyzing changes in neural features during stimulation periods. Since the decoder operates only on neural features, and not on the raw recordings, artifacts can only reduce decoder performance via their effect on neural features. Neural feature changes were analyzed at two different time scales: first, at the sample-by-sample resolution during artifact periods, and second by averaging features in 20ms windows (representing typical iBCI systems).

To assess of the effect of stimulation artifacts on feature computation, threshold crossings (TX) and high frequency spike power (HFSP) were calculated at each timestep and the signals were aligned to stimulation onset. Baseline feature values (calculated as the average value during non-artifact periods) were then subtracted to obtain the average change in features during stimulation for each artifact reduction method. Figure 6(A & B) shows these traces for one representative electrode during a block of surface and intramuscular stimulation. If the artifacts were completely eliminated during cleaning, we would expect a flat trace at 0 for both features, indicating no average increase during stimulation. Both surface and intramuscular stimulation resulted in high feature values during stimulation periods (left columns), aligned in time with the average stimulation artifact. Cleaning the signal using CAR reduced the average artifact and the TX and HFSP features. However, this method was not adequate for surface stimulation, which had large increases in both features during stimulation. LRR performed much better, reducing the average artifact and the added feature values during both stimulation types. The aligned features after LRR cleaning were much smaller for all conditions, showing that LRR performed well at reducing artifacts and restoring normal feature values during muscle stimulation.

We also explored feature changes on a larger time scale, binning features into 20ms windows typical of common iBCI feature extraction. The feature changes were quantified by calculating the average value of bins containing stimulation and normalizing to the average value of bins without stimulation. Average feature values were aggregated for all channels (192) and all blocks (15 intramuscular stimulation, 7 surface stimulation) and summarized in figure 6(C & D). During both stimulation types, feature means increased in bins containing artifacts (black bars higher than baseline). CAR reduced these feature changes minimally for surface stimulation, but almost completely eliminated them for intramuscular artifacts ($1.04\times$ higher than baseline). LRR outperformed CAR for both stimulation conditions (orange bars lower than blue and black), but still had moderate increases in TX rate during surface stimulation. These results suggest that TX features may be more affected by residual artifacts than HFSP features. Additionally, LRR median features values were quite low, indicating generally good performance on most channels and blocks. However, mean LRR feature values were considerably higher, representing an unequal contribution or error across channels (some channels had large feature mean increases while most had very small changes). These highest contributing channels tended to have very small baseline values, which could result in larger feature mean increases for the same residual artifact. As expected, blanked data presented very similar feature means between stimulation and non-stimulation bins, indicating that the duration of blanking was adequate to exclude the whole artifact for both stimulation conditions.

We examined how stimulation artifacts affected our TX features by comparing the neural signal snippets that triggered threshold crossings during non-stimulation and stimulation periods (supplementary figure 3). The LRR and CAR waveforms during non-stimulation periods were similar ($r = 0.99$), suggesting that LRR does not distort normal spike waveforms compared to the standard CAR. During stimulation periods, the average CAR snippet was strongly distorted compared to the average non-stimulation period spike waveform ($r = 0.54$), but the average LRR snippet was nearly identical to non-stimulation period waveforms ($r = 0.99$). This suggests the LRR cleaned signal had a much smaller

number of artifact-triggered TX events compared to the CAR cleaned signal, an indicator of good artifact cleaning performance.

Decoding Performance

Finally, we directly investigated the impact of stimulation artifacts on actual iBCI performance. To do so, separate decoders predicting 3D movements of the virtual arm were built for each artifact reduction technique using non-stimulation data, and then applied to blocks with stimulation (figure 2). iBCI performance was measured by computing the SNR of the decoder during periods with stimulation, and normalizing to its performance to typical system baseline (non-stimulation bins with CAR applied). Two analyses were performed, first using only artifact period data (figure 7(A)) to understand the information retained after artifact reduction, and second using larger bins of mixed clean and artifact data to understand typical system performance (figure 7(B)).

Results show that significant movement-related information can be recovered from artifact periods after cleaning (figure 7(A)), with LRR and CAR achieving performance during intramuscular stimulation not significantly different from baseline (adjusted $p > 0.05$), and LRR nearly recovering (90%) full performance during surface stimulation. When combining artifact and non-artifact data, again CAR and LRR achieved decoder SNR not significantly different from baseline (adjusted $p > 0.05$) during intramuscular stimulation, but only LRR recovered >90% SNR during surface stimulation. Blanking performed moderately well during surface stimulation (84%), but during intramuscular stimulation it actually reduced SNR more than the raw artifacts.

Additional offline experiments were performed to determine how the amount of blanking affects decoding performance. While artifact reduction techniques were able to recover informative activity in our FES+iBCI system, this may not always be the case (if stimulation saturates recording amplifiers for example). Therefore, it is helpful to fully understand how blanking duration will affect information loss when designing future systems. We started with 10 blocks of artifact-free data and applied increasing durations of blanking from 0 to 19.5ms to every 20ms bin in an offline decoding analysis. The results were normalized per block to the decoder SNR when no blanking was applied. As expected, the normalized decoder SNR decreased as blanking duration was increased (figure 8). We derived a model for the reduction in SNR as a function of blanking duration, assuming decoder noise is normally distributed, uncorrelated, and independent. From these assumptions, it follows that the blanked decoder SNR, normalized to a full bin width SNR, is

$SNR_{Blanked} = SNR(\sqrt{1 - b})$, where b is the blanking coefficient (fraction of data that is blanked, 0 representing no blanking and 1 representing complete blanking) (See supplement section 1 for full derivation). This model accurately fit the mean of our offline results ($R^2 > .98$), and can be used to predict the SNR reduction for arbitrary blanking durations and frequencies.

4. Discussion

Overview

In this study, we characterized peripheral electrical stimulation artifacts (from both surface and intramuscular electrodes) that contaminated the neural recordings of an FES+iBCI system and tested several different techniques for reducing them. Using data from one participant of the BrainGate2 clinical trial, who had two 96-channel intracortical microelectrode arrays and 36 intramuscular FES electrodes implanted, we offline compared three artifact reduction techniques (common average reference (CAR), linear regression reference (LRR), and blanking). The predominant finding of the study was that significant performance improvements could be achieved using the LRR artifact reduction technique compared to blanking or CAR. We also demonstrated large reductions in artifact magnitudes, and used decoding performance results to show that informative neural activity could be recovered during artifact periods. Finally, we developed a model predicting iBCI performance degradation during blanking.

Recovering informative activity from artifact periods

Informative neural activity could be recovered from contaminated data (figure 7(A)), even during surface stimulation, when artifacts are orders of magnitude larger than the underlying neural information (figure 5(B)). However, extrapolating these results to real iBCI system performance is complicated by the fact that most systems will stimulate and bin neural data at different frequencies, leading to a mixture of clean and contaminated data. Therefore, we performed a second analysis using typical feature bin durations, which contained a mixture of artifact and clean data (figure 7(B)). Both analyses showed that artifact reduction methods can outperform blanking, indicating that useful activity exists underneath the artifacts that blanking needlessly excludes.

While informative, the data presented does not represent an upper limit on potential performance. Increased decoding accuracy could be achieved by carefully separating the artifact and clean data epochs, performing decoding separately, and combining the predictions based on the average decoding quality of each epoch. Perhaps more simply, an alternative strategy could apply a single decoder to mixed-bin data (as in figure 7(B)), but with decoder predictions from artifact-contaminated bins down-weighted based on the amount of artifact. While the artifacts in our system did not severely limit performance, these strategies may be necessary for alternative applications with higher amplitude, longer duration, or more frequent artifacts.

Restored performance after FES+iBCI artifact reduction

Our decoding results showed that unmitigated stimulation artifacts would significantly reduce iBCI performance (figure 7). One might have initially expected this noise to put a hard limit on FES+iBCI system performance. However, normal decoding performance was recovered by applying artifact reduction techniques offline using typical system hardware and feature extraction processes. The LRR method achieved >90% baseline performance during surface stimulation and both CAR and LRR restored full performance during intramuscular stimulation. These data show that simple reduction techniques can reduce

large-amplitude stimulation artifacts and enable more accurate movement decoding than blanking. While our study compared artifact reduction techniques offline, our methods can be easily implemented for online use of closed-loop FES+iBCI control with few modifications to existing systems or procedures. Therefore, stimulation artifacts should not be considered a performance-limiting factor in FES+iBCI systems.

Predicting performance degradation due to blanking

We built a model that accurately predicts the performance reduction that results from various durations of blanking (figure 8). The model shows a non-linear drop in performance, with only small decreases in decoder SNR for short blanking durations. These results match our understanding of the diminishing returns of averaging out noise; for normally distributed

noise, we expect a $\frac{1}{\sqrt{n}}$ reduction in the standard deviation of the noise (where n is the number of samples being averaging). When averaging across many samples, there are diminishing returns in the amount of information extracted from the neural activity, so removing small portions does not affect performance as much as a linear model might assume. Therefore, blanking may be an acceptable artifact reduction technique for applications with short stimulation periods or relatively lenient accuracy requirements.

For simplicity, our blanking model assumed equal blanking durations at every timestep. However, systems may have unmatched stimulation and recording frequencies. For example, our system stimulates at 12.5Hz but bins neural data at 50Hz. Thus, our offline decoding results do not match those predicted by the model, where blanking appears to degrade our performance much more than the model predicts. This difference is due to the fact that performance is substantially reduced when the blanked bins (which are noisier because they contain less data after blanking) are treated the same as the fully-clean bins by the decoder. Generally, to achieve maximal performance, the decoder output of bins that contain less information must be down-weighted (scalar magnitude reduction). Our model can be useful for determining the amount of reduction that should be applied, because it predicts the decoding noise based on blanking duration. If down-weighting is not used, one should expect a larger performance degradation due to blanking than that shown in figure 8.

Our model and results could inform the design of novel stimulation paradigms and recording systems, helping to choose the best architecture and artifact rejection technique needed to achieve a desired performance level.

Advantages of linear regression reference

There was a strong correlation in artifact waveforms across channels on the microelectrode arrays (figure 3(C)), which suggested that artifact reduction methods that utilize data from all electrodes within an array might perform better than alternative methods (filtering, template subtraction, curve fitting, etc.). We compared LRR, which is trained on artifact data to optimally weight each channels' contribution to the reference, to the standard CAR. The LRR weights typically were much higher for nearby channels, possibly drawing comparisons to more localized referencing strategies such as a Laplacian [34]. However, initial results implementing a small Laplacian reference yielded artifact magnitude

reductions that were similar to CAR and much worse than those achieved by LRR (supplementary figures 4-5). Additionally, we saw large variability in weight patterns across channels (figure 4), highlighting the advantage of LRR over CAR (or any similar fixed-referencing strategy) by producing channel-specific references built with knowledge of the actual artifact waveforms on all channels. We trained a single set of LRR weights for each block, during which the limb was manipulated through an extensive range of joint and limb positions. Restoring normal decoding performance during this task highlights the ability of LRR to generalize across many stimulation parameters and limb positions for both intramuscular and surface stimulation. The weights were largely stable in time (despite changing stimulation parameters), suggesting that minimal re-training may be necessary. Additionally, there are many LRR parameters which could be optimized (amount of training data, pre/post filtering, etc.) to potentially achieve even better performance. Another advantage of the LRR method is its adaptability to greater stimulation complexity (e.g. increased number of channels, longer stimulation durations, etc.), which may be especially important in sensory feedback applications [35]. Whereas blanking performance would continuously decrease (figure 8), LRR and other artifact reduction techniques would likely still recover useful information from those longer stimulation periods. Since the LRR method achieved the best performance in our tests (figures 5–7), it should prove useful for future FES+iBCI systems whenever CAR is not sufficient.

Artifact reduction in biomedical context

We compared two alternative FES paradigms that have been presented in recent studies: surface [9] and intramuscular [10] stimulation. Using typical stimulation parameters, our results showed that surface stimulation created much larger artifacts than intramuscular stimulation (figure 5(B)). However, after applying LRR the decoding performance during surface stimulation artifact periods was only ~10% worse than non-stimulation periods, indicating that good performance can still be achieved despite the large artifact. In the case of intramuscular stimulation, artifacts were much smaller and full performance recovery could be achieved with the simpler CAR method. Thus, adequate performance can be achieved under both the intramuscular and surface stimulation paradigms, but intramuscular stimulation has the advantage of achieving nearly complete artifact reduction with simpler signal processing (CAR). The LRR artifact reduction technique performed better than blanking in both stimulation conditions, suggesting that in similar applications, blanking should be avoided when trying to maximize performance.

FES+iBCI systems are just one of a growing number of rehabilitation and research systems that utilize simultaneous stimulation and recording. Recent work has been reported on closed-loop deep-brain-stimulation [36], intracortical microstimulation and recording [17,19,35], among others. This study characterizes many of the stimulation artifact properties that are often underreported in the literature and addresses some of the perceived hurdles of a closed-loop system. Our system exhibited some properties that aided successful artifact reduction, namely that we recorded from many channels with highly correlated artifacts. Other systems with similar properties (such as intracortical microstimulation) may benefit from implementing the LRR method for artifact reduction. In cases where stimulation artifacts saturate the recording amplifiers (preventing recovery of underlying

information), our model of performance degradation based on blanking duration could be useful for system designers.

5. Conclusion

Recent advances in rehabilitation technologies have restored intention-controlled functional movement to paralyzed people using combined FES+iBCI systems [9,10]. However, electrical stimulation of the muscles generates large artifacts on the intracortical neural recordings, corrupting the signals and degrading system performance. This study implemented an artifact reduction technique, linear regression reference (LRR), which outperformed blanking and common average reference and restored >90% normal system performance. By taking advantage of correlations in noise across many recording electrodes, LRR recovered informative neural activity during artifact periods. This technique is potentially useful for noise reduction in a variety of neuroprosthetic applications with multiple recording electrodes (electrocorticography, closed-loop deep brain stimulation, intracortical microstimulation, etc.). These results show that stimulation artifacts during either surface or intramuscular stimulation can be significantly reduced such that the artifacts are not a performance-limiting factor in FES+iBCI systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank participant T8 for his time and continuous devotion to our studies. Support for this study was provided primarily by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) of the National Institutes of Health (NIH) under award number R01HD077220, and by NICHD-NIH N01HD53403, the National Institute on Deafness and Other Communication Disorders of NIH under award number R01DC009899, and by T32 funding from the National Institute of Biomedical Imaging and Bioengineering of the NIH (5T32EB004314-15). Support was also provided by the United States Department of Veterans Affairs, Rehabilitation Research and Development Service, under award numbers B4853C, B6453R, and N9228C, and by the MGH-Deane Institute, The Executive Committee on Research (ECOR) of Massachusetts General Hospital.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, or the Department of Veterans Affairs or the United States Government. CAUTION: Investigational Device. Limited by Federal Law to Investigational Use.

References

1. Peckham PH, Keith MW, Freehafer AA. Restoration of Functional Control by Electrical Stimulation in the Upper Extremity of the Quadriplegic Patient. *J Bone Jt Surg.* 1988; 70:144–8.
2. Memberg WD, Polasek KH, Hart RL, Bryden AM, Otr L, Kilgore KL, Nemunaitis GA, Huyen HA, Keith MW, Kirsch RF. Implanted Neuroprosthesis for Restoring Arm and Hand Function in People With High Level Tetraplegia. *Arch Phys Med Rehabil.* 2014; 95
3. Simeral JD, Kim S-P, Black MJ, Donoghue JP, Hochberg LR. Neural control of cursor trajectory and click by a human with tetraplegia 1000 days after implant of an intracortical microelectrode array. *J Neural Eng.* 2011; 8:25027.
4. Bacher D, Jarosiewicz B, Masse NY, Stavisky SD, Simeral JD, Newell K, Oakley EM, Cash SS, Friebs G, Hochberg LR. Neural Point-and-Click Communication by a Person With Incomplete Locked-In Syndrome. *Neurorehabil Neural Repair.* 2015; 29:462–71. [PubMed: 25385765]

5. Pandarinath C, Nuyujukian P, Blabe CH, Sorice BL, Saab J, Willett FR, Hochberg LR, Shenoy KV, Henderson JM. High performance communication by people with paralysis using an intracortical brain-computer interface. *Elife*. 2017; 6:1–27.
6. Collinger JL, Wodlinger B, Downey JE, Wang W, Tyler-Kabara EC, Weber DJ, McMorland AJ, Velliste M, Boninger ML, Schwartz AB. Supplementary appendix. *Lancet*. 2012; 6736
7. Hochberg LR, Bacher D, Jarosiewicz B, Masse NY, Simeral JD, Vogel J, Haddadin S, Liu J, Cash SS, van der Smagt P, Donoghue JP. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*. 2012; 485:372–5. [PubMed: 22596161]
8. Hochberg LR, Serruya MD, Friehs GM, Mukand J a, Saleh M, Caplan AH, Branner A, Chen D, Penn RD, Donoghue JP. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*. 2006; 442:164–71. [PubMed: 16838014]
9. Bouton CE, Shaikhouni A, Annetta NV, Bockbrader MA, Friedenberg DA, Nielson DM, Sharma G, Sederberg PB, Glenn BC, Mysiw WJ, Morgan AG, Deogaonkar M, Rezai AR. Restoring cortical control of functional movement in a human with quadriplegia. *Nature*. 2016; 0:1–13.
10. Ajiboye AB, Willett FR, Young DR, Memberg W, Walters BC, Sweet JA, Hoyer HA, Keith MW, Peckham PH, Simeral JD, Donoghue JP, Miller JP, Hochberg LR, Kirsch RF. Restoration of reaching and grasping in a person with tetraplegia through brain-controlled muscle stimulation: a proof-of-concept demonstration. *Lancet*. 2017; 6736 (In press).
11. Hashimoto T, Elder CM, Vitek JL. A template subtraction method for stimulus artifact removal in high-frequency deep brain stimulation. *J Neurosci Methods*. 2002; 113:181–6. [PubMed: 11772439]
12. Hoffmann U, Cho W, Ramos-Murguialday A, Keller T. Detection and removal of stimulation artifacts in electroencephalogram recordings. *Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS*. 2011:7159–62.
13. O’Keeffe DT, Lyons GM, Donnelly AE, Byrne CA. Stimulus artifact removal using a software-based two-stage peak detection algorithm. *J Neurosci Methods*. 2001; 109:137–45. [PubMed: 11513948]
14. Gnadt JW, Echols SD, Yildirim A, Zhang H, Paul K. Spectral cancellation of microstimulation artifact for simultaneous neural recording in situ. *IEEE Trans Biomed Eng*. 2003; 50:1129–35. [PubMed: 14560765]
15. Rossi L, Foffani G, Marceglia S, Bracchi F, Barbieri S, Priori A. An electronic device for artefact suppression in human local field potential recordings during deep brain stimulation. *J Neural Eng*. 2007; 4:96–106. [PubMed: 17409484]
16. Kent AR, Grill WM. Recording evoked potentials during deep brain stimulation: development and validation of instrumentation to suppress the stimulus artefact. *J Neural Eng*. 2012; 9:36004.
17. O’Doherty JE, Lebedev MA, Ifft PJ, Zhuang KZ, Shokur S, Bleuler H, Nicolelis MAL. Active tactile exploration using a brain–machine–brain interface. *Nature*. 2011; 479:228–31. [PubMed: 21976021]
18. Freeman JA. An electronic stimulus artifact suppressor. *Electroencephalogr Clin Neurophysiol*. 1971; 31:170–2. [PubMed: 4104707]
19. O’Doherty JE, Lebedev MA, Zheng Li, Nicolelis MAL, Li Z, Nicolelis MAL. Virtual active touch using randomly patterned intracortical microstimulation. *IEEE Trans Neural Syst Rehabil Eng*. 2012; 20:85–93. [PubMed: 22207642]
20. Hines AE, Crago PE, Chapman GJ, Billian C. Stimulus artifact removal in EMG from muscles adjacent to stimulated muscles. *J Neurosci Methods*. 1996; 64:55–62. [PubMed: 8869484]
21. Toleikis RJ, Metman LV, Pilitsis JG, Barborica A, Toleikis SC, Bakay RAE. Effect of intraoperative subthalamic nucleus DBS on human single-unit activity in the ipsilateral and contralateral subthalamic nucleus. *J Neurosurg*. 2012; 116:1134–43. [PubMed: 22339160]
22. Heffer LF, Fallon JB. A novel stimulus artifact removal technique for high-rate electrical stimulation. *J Neurosci Methods*. 2008; 170:277–84. [PubMed: 18339428]
23. Walter A, Ramos Murguialday A, Spüler M, Naros G, Leão MT, Gharabaghi A, Rosenstiel W, Birbaumer N, Bogdan M. Coupling BCI and cortical stimulation for brain-state-dependent stimulation: methods for spectral estimation in the presence of stimulation after-effects. *Front Neural Circuits*. 2012; 6:87. [PubMed: 23162436]

24. Litvak LM, Smith ZM, Delgutte B, Eddington DK. Desynchronization of electrically evoked auditory-nerve activity by high-frequency pulse trains of long duration. *J Acoust Soc Am*. 2003; 114:2066–78. [PubMed: 14587606]
25. Klaes C, Shi Y, Kellis S, Minxha J, Revechkis B, Andersen RA. A cognitive neuroprosthetic that uses cortical stimulation for somatosensory feedback. *J Neural Eng*. 2014; 11:56024.
26. Wagenaar DA, Potter SM. Real-time multi-channel stimulus artifact suppression by local curve fitting. *J Neurosci Methods*. 2002; 120:113–20. [PubMed: 12385761]
27. Anderson ME, Postupna N, Ruffo M, Marjorie E, Postupna N, Effects MR. Effects of high-frequency stimulation in the internal globus pallidus on the activity of thalamic neurons in the awake monkey. *J Neurophysiol*. 2003; 89:1150–60. [PubMed: 12574488]
28. Collinger JL, Wodlinger B, Downey JE, Wang W, Tyler-Kabara EC, Weber DJ, McMorland AJ, Velliste M, Boninger ML, Schwartz AB. High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet*. 2012; 6736
29. Jarosiewicz B, Sarma AA, Bacher D, Masse NY, Simeral JD, Sorice B, Oakley EM, Blabe C, Pandarinath C, Gilja V, Cash SS, Eskandar EN, Friehs G, Henderson JM, Shenoy KV, Donoghue JP, Hochberg LR. Virtual typing by people with tetraplegia using a self-calibrating intracortical brain-computer interface. *Sci Transl Med*. 2015; 7
30. Velliste M, Perel S, Spalding MC, Whitford AS, Schwartz AB. Cortical control of a prosthetic arm for self-feeding. *Nature*. 2008; 453:1098–101. [PubMed: 18509337]
31. Musial P, Baker S, Gerstein G, King E, Keating J. Signal-to-noise ratio improvement in multiple electrode recording. *J Neurosci Methods*. 2002; 115:29–43. [PubMed: 11897361]
32. Willett FR, Murphy BA, Memberg WD, Blabe CH, Pandarinath C, Walter BL, Sweet JA, Miller JP, Henderson JM, Shenoy KV, Hochberg LR, Kirsch RF, Ajiboye AB. Signal-independent noise in intracortical brain–computer interfaces causes movement time properties inconsistent with Fitts’ law. *J Neural Eng*. 2017; 14:26010.
33. Willett FR, Pandarinath C, Jarosiewicz B, Murphy BA, Memberg WD, Blabe CH, Saab J, Walter BL, Sweet JA, Miller JP, Henderson JM, Shenoy KV, Simeral JD, Hochberg LR, Kirsch RF, Ajiboye AB. Feedback control policies employed by people using intracortical brain–computer interfaces. *J Neural Eng*. 2017; 14:16001.
34. McFarland DJ, McCane LM, David SV, Wolpaw JR. Spatial filter selection for EEG-based communication. *Electroencephalogr Clin Neurophysiol*. 1997; 103:386–94. [PubMed: 9305287]
35. Flesher, S., Downey, J., Collinger, J., Foldes, S., Weiss, J., Tyler-Kabara, E., Bensmaia, S., Schwartz, A., Boninger, M., Gaunt, R. Intracortical Microstimulation as a Feedback Source for Brain-Computer Interface Users. In: Guger, C., Allison, B., Lebedev, MA., editors. *Brain-Computer Interface Research*. Cham, Switzerland: SpringerBriefs in Electrical and Computer Engineering; 2017. p. 43–54.
36. Little S, Pogosyan A, Neal S, Zavala B, Zrinzo L, Hariz M, Foltynie T, Limousin P, Ashkan K, FitzGerald J, Green AL, Aziz TZ, Brown P. Adaptive deep brain stimulation in advanced Parkinson disease. *Ann Neurol*. 2013; 74:449–57. [PubMed: 23852650]

Significance

The LRR method was effective at reducing electrical artifacts resulting from both intramuscular and surface FES, and almost completely restored iBCI decoding performance (>90% recovery for surface stimulation and full recovery for intramuscular stimulation). The results demonstrate that FES-induced artifacts can be easily mitigated in FES+iBCI systems by using LRR for artifact reduction, and suggest that the LRR method may also be useful in other noise reduction applications.

Figure 1.

System diagram. (A) The study participant had two intracortical microelectrode arrays implanted in the motor cortex, as well as 32 stimulating intramuscular electrodes and four anode intramuscular electrodes with percutaneous connections. During functional electrical stimulation (FES), electric fields generated by the peripheral muscle stimulation extended to the cortex, contaminating the neural recordings. (B) Pictures of [Left] implantation site of the intracortical arrays, [Middle] intramuscular stimulating electrodes, and [Right] temporary surface patch electrodes used in this study.

Figure 2.

Offline decoding methods and metrics. (A) Block diagram of the methods for comparing offline decoding performance between artifact reduction methods. Starting with two sets of neural data, non-stimulation blocks (top) used for decoder training and blocks with stimulation (bottom) used for performance testing, we calculate feature vectors ($F_{384 \times t}$) for each of our artifact reduction techniques. We use the non-stimulation features to build ‘clean’ condition-specific decoders (top). We apply those ‘clean’ decoders to the artifact-reduced features for each technique (bottom) and measure the performance through decoder signal-to-noise ratio (SNR). (B) Illustration of the decoder SNR metric used to quantify offline decoding performance. Noise (e_d) was defined as the difference between decoded velocity (u_d) and empirically modeled intended velocity (c_d), while signal was defined as the average magnitude of the decoded velocity when the cursor was far from the target. SNR was defined as the ratio of the signal to the standard deviation of the noise.

Figure 3.

Artifact characterization. (A) Recordings from a representative microelectrode channel during single electrode functional electrical stimulation (FES). [Top] recordings made during stimulation via surface patch electrodes placed on the deltoid, stimulated at 12.5Hz with 60mA. [Bottom] Recordings made during stimulation via an intramuscular electrode targeting the same muscle, stimulated at 12.5Hz with 20mA. (B) Stimulation artifact waveforms recorded for each stimulating-electrode type, averaged over 1000+ trials (grey region denotes the standard deviation). The peak-to-peak magnitudes of the waveforms are indicated and compared to typical background activity magnitudes ($110\mu V_{pp}$). (C) Median correlation between trial-averaged stimulation artifact waveforms (baseline activity $< 1 \mu V_{pp}$) between channels within each array.

Figure 4.

Examples of linear regression reference (LRR) weights from the medial array. Single channel examples of LRR weights, showing representative examples (Ch. 136, 125, 170) and less typical examples (Ch. 158, 148). Each image shows LRR weights for the channel indicated by the red X. Each colored pixel represents the magnitude of the coefficient for the channel in that physical location on the array. All images show average coefficients across all blocks of surface or intramuscular stimulation.

Figure 5.

Artifact reduction performance (A) Artifact magnitude reduction illustrated for a representative channel. [Top] Neural recordings made during surface stimulation after artifact reduction and bandpass filtering. Note the larger y-axis in the raw subplot, and the reduced artifact amplitudes after artifact reduction. The blanked trace removes 5ms of data during stimulation. [Bottom] Averages of 1000+ stimulation events for each reduction technique. During blanking, no signal is shown. (B) Distribution of the artifact magnitudes for each stimulation and artifact reduction condition during single-electrode stimulation blocks. Data includes one point per channel (192 total) per block (6 surface and 4 intramuscular blocks). Colored bars represent the interquartile range, horizontal lines represent the median, colored dots represents the mean, and dotted lines extend to the maximum and minimum non-outlier values (defined as points outside $1.5\times$ the inter-quartile

range). (C) Neural frequency content after artifact reduction as quantified by power spectral density curves, calculated on the continuous recordings before (Raw) and after CAR and LRR cleaning. PSD curves were averaged across all channels on one array for one non-stimulation block and three blocks each of surface and intramuscular stimulation. The grey background represents spike band frequency range (250-5000Hz).

Figure 6.

Effect of artifact reduction on neural features. (A & B) [Top Row] Artifact waveform for a representative channel during one surface block (A) or intramuscular block (B), computed by averaging the recorded voltage over 1000+ stimulation events. [Middle Row] Average number of threshold crossing events (TX) counted at each sample of the 15kHz signal, with baseline values (during non-stimulation periods) subtracted. [Bottom Row] Aligned and averaged HFSP features during artifact periods with baseline values subtracted. Note the different y-axis values across conditions for the artifact magnitude and HFSP plots. (C & D) Binned feature changes during stimulation. Aggregated across all blocks and channels, boxplots show distributions of average feature values during 20ms bins containing stimulation artifact, normalized per-block and per-channel to the average feature value without stimulation. Colored bars represent the interquartile range, horizontal lines represent the median, colored dots represent the mean, dotted lines extend to the maximum and

minimum non-outlier values (defined as points outside 1.5x the inter-quartile range). Data are separated for each stimulation type (column) and cleaning method (color) for threshold crossings (C) and high-frequency spike power (D). Features with low firing rates (<5 crossings/second) were excluded from TX distributions.

Figure 7.

Decoding performance during artifact periods. Signal-to-noise ratio (SNR) was normalized within each block by dividing by the SNR of a decoder applied to clean data snippets of the same duration. The dots represent mean and bars extend to 95% confidence interval. (A) Decoder SNR using only data from artifact periods (5ms surface, 8ms intramuscular). (B) Decoder SNR on mixed-bin data (unaligned 20ms bins containing at least partial stimulation artifact).

Figure 8.

Modeling how decoder SNR declines as blanking is applied. Decoder SNR is shown as a function of blanking coefficient, computed offline. The blanking coefficient represents the fraction of blanked data (0 representing no blanking, 1 representing complete blanking). Ten blocks of data are shown as light grey traces and the model prediction is shown in black. The model is derived in Supplementary Section 1. The model predicts how SNR decreases with the square root of the blanking duration, and closely matches the real data.

Table 1

Artifact Reduction Techniques. Description of the three artifact reduction techniques explored in this paper, equations describing their application, and their primary assumptions about the stimulation artifacts. R_c is a vector of raw recorded microelectrode recordings on channel c , S_c is the cleaned signal from channel c , CR is the set of channels included in the common average reference, n is the number of channels in CR (80), $\vec{w}_c$ is a vector of LRR weight coefficients calculated specifically for channel c , and X is a matrix of raw recordings for all channels in an array except c (where each row contains the recordings for a single time step).

Cleaning Method	Application	Assumption
Blanking	$S_c = \begin{cases} R_c & \text{:stim off} \\ \text{simulated data} & \text{:stim on} \end{cases}$	Entire corrupted signal should be excluded
Common Average Reference (CAR)	$S_c = R_c - \frac{1}{n} \sum_{i \in CR} S_i$	Artifact is similar across all channels, and can be predicted by a fixed average
Linear Regression Reference (LRR)	$S_c = R_c - X \vec{w}_c$	Artifact is similar across some channels and can be predicted by a weighted sum of other channels