Journal of Neuroscience Methods 191 (2010) 45–59

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Generalized framework for stimulus artifact removal

Yaara Erez [a], Hadass Tischler [a], Anan Moran [a], Izhar Bar-Gad [a,b,*]

[a] Gonda Brain Research Center, Bar Ilan University, Ramat Gan, Israel
[b] Goodman Faculty of Life Sciences, Bar Ilan University, Ramat Gan, Israel

ARTICLE INFO

Article history:
Received 8 April 2010
Received in revised form 30 May 2010
Accepted 2 June 2010

Keywords:
Electrical stimulation
Magnetic stimulation
Stimulation artifact
Artifact removal
Extracellular recording
SARGE

ABSTRACT

Stimulation is extensively used in neuroscience research in diverse fields ranging from cognitive to clinical. Studying the effect of electrical and magnetic stimulation on neuronal activity is complicated by large stimulation-derived artifacts on the recording electrodes, which mask the spiking activity. Multiple studies have suggested a variety of solutions for the removal of artifacts and were typically directed at specific stimulation setups. In this study we introduce a generalized framework for stimulus artifacts removal, the Stimulus Artifact Removal Graphical Environment (SARGE). The framework provides an encapsulated environment for a multi-stage removal process, starting from the stimulus pulse detection, through estimation of the artifacts and their removal, and finally to signal reconstruction and the assessment of removal quality. The framework provides the user with subjective graphical and objective quantitative tools for assessing the resulting signal, and the ability to adjust the process to optimize the results. This extendable publicly available framework supports different types of stimulation, stimulation patterns and shapes, and a variety of artifact estimation methods. We exemplify the removal of artifacts generated by electrical micro- and macro-stimulation and magnetic stimulation and different stimulation protocols. The use of different estimation methods, such as averaging and function fitting is demonstrated, and the differences between them are discussed. Finally, the quality of removal is assessed and validated using quantitative measures and combined experimental-simulation studies. The framework marks a shift from "algorithm" and "data" centric approach to a "workflow" centric approach, thus introducing an innovative concept to the artifact removal process.

1. Introduction

Stimulation in different areas of the central nervous system is widely used in neuroscience, for both exploring basic science, and for treating the symptoms of multiple pathological states (Benabid et al., 1994; Wichmann and DeLong, 2006). Stimulation is carried out in a large spectrum of experiments including in vitro (Wagenaar et al., 2005), in vivo, and even human studies, and may take different forms ranging from electrical micro-stimulation (Dostrovsky et al., 2000; Erez et al., 2009) to electrical macro-stimulation (Hashimoto et al., 2003; Carlson et al., 2010) and magnetic stimulation (Moliadze et al., 2003; Strafella et al., 2004). The study of neuronal activity during electrical or magnetic stimulation is a challenge because of the large artifacts generated by the stimulation pulses which mask the spiking activity and make standard methods of spike identification and sorting impractical. When a stimulus pulse is delivered, large electrical transients are evident in the signal recorded by the electrodes. These large, rapid electrical changes, which in most cases reach saturation in the acquisition systems, result in a period during which spikes cannot be extracted at all. Following this initial period, the stimulus effect, termed 'stimulus artifact', is still present for a time period of milliseconds or even tens or hundreds of milliseconds, distorting the recorded signal (Wagenaar and Potter, 2002). These changes are a result of the combination of several factors such as the capacitive properties of the acquisition device, the tissue properties, the electrode properties, and others. During this second period spikes are distorted but may be observed, although their identification in the raw recorded signal by automatic or even semi-automatic methods is impossible in most cases. These stimulation artifacts can be partially suppressed by online hardware and software solutions (Wagenaar and Potter, 2002; Brown et al., 2008). However when these are not applicable, or do not fully remove the artifacts, the use of offline artifact removal algorithms is a necessity.

The shape of a stimulus artifact is influenced by a complex nonlinear combination of multiple parameters, such as the stimulation type (e.g., electrical, magnetic), properties of the stimulating and recording electrodes (e.g., impedance, shape of tip), the distance between the stimulation and recording sites, single stimulus pulse properties (mono- vs. biphasic, duration and amplitude), the stimulation protocol (such as frequency, temporal organization) and

* Corresponding author at: Gonda Brain Research Center, Bar Ilan University, Ramat Gan 52900, Israel. Tel.: +972 3 5317141.
E-mail address: bargadi@mail.biu.ac.il (I. Bar-Gad).

0165-0270/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneumeth.2010.06.005

even properties of the amplification, filtering and acquisition system (Ranck, Jr., 1975; Merrill et al., 2005). Previous studies have demonstrated the utility of offline removal of stimulation artifacts using different methods, which can be divided into three main types. The first type includes methods in which the stimulus artifact is removed, including its decaying tail, and set to baseline values or replaced by a linear interpolation linking the two closest remaining sampled points before and after the pulse (O'Keeffe et al., 2001; Heffer and Fallon, 2008). This type of removal is efficient in terms of time, but inevitably leads to a loss of spiking activity within the whole tail of the artifact. The other two types try to cope with the tail of the artifact to preserve the spiking information that might exist there. The second type covers averaging-based methods. These methods assume that consecutive artifacts of stimulation pulses are similar in shape, and therefore can be removed using the average shape (Wichmann, 2000; Hashimoto et al., 2002; Montgomery, Jr. et al., 2005), while preserving the spiking activity which differs following different stimulation pulses. The third type includes function fitting methods, in which a specific function (or function family) is fitted for each stimulus artifact. After the subtraction of the function, which approximates the slow components of change in the signal, the spiking activity, which is characterized by rapid changes in the signal, can be extracted. Different functions may be used for this artifact removal, and are primarily based on their resemblance to the shape of the artifact. For example, in many cases the artifacts decay exponentially; thus, an exponential function, or the sum of two exponential functions, may be fitted to the artifact (Harding, 1991). Another family of functions that may be fitted to the artifacts is polynomials of varying degrees, where a single polynomial can be fitted for the whole artifact, or a different polynomial can be fitted for a predefined window around each sampled point (Wagenaar and Potter, 2002).

Previous solutions for the removal of artifacts focused on specific data sets of stimulation which were performed for special purposes. This required the *de novo* development of an artifact removal solution tailor-made for each stimulation configuration. In this paper we address this problem by introducing a general framework, the Stimulus Artifact Removal Graphical Environment (SARGE), aimed at the removal of artifacts from multiple types of stimulations (micro-stimulation, macro-stimulation, magnetic stimulation) and stimulation patterns (high frequency, low frequency, etc.). The SARGE has a graphical interface for a multi-stage process that considers multiple aspects of the stimulus and recorded signal to best remove the artifacts and reconstruct the signal, thus enabling the extraction of spiking activity. It combines semi-automatic calculations, but also makes it possible to adjust various parameters to different artifacts. The framework provides a convenient graphical user interface (GUI) and workspace in which large data sets can be relatively easily processed, as increasing amounts of data can be acquired as a result of the advanced technology. Additionally, quality assessment tools are provided within the framework to insure accurate results and successful extraction of spiking activity in later stages of analysis. In this study, we present the SARGE and the methods of analysis, and illustrate their application on simulated data and on experimental recordings from behaving animals.

2. Materials and methods

2.1. Animals and electrophysiological recordings

Neurophysiological recordings from three cynomolgus (*Macaca fascicularis*) male monkeys were used. The monkeys' water and food consumption and weight were checked daily and their health was monitored by a veterinarian. All procedures were in accordance with the *National Institutes of Health Guide for the Care and Use of Laboratory Animals* and *Bar Ilan University guidelines for the use and care of laboratory animals in research* and were approved and supervised by the *Institutional Animal Care and Use Committee (IACUC)*. Full details of the experimental protocol appear elsewhere (Erez et al., 2009). Briefly, data were acquired simultaneously via up to 12 glass-coated tungsten microelectrodes (impedance 0.2–0.7 MΩ at 1 kHz). The electrode signal was amplified (*1000) and band-pass filtered (2–8000 Hz four-pole Butterworth filter) (MCP-Plus 4.10, Alpha–Omega Engineering, Nazareth, Israel). The signal was continuously sampled at 40 kHz with 14-bit resolution (Alphamap 10.10, Alpha–Omega Engineering), yielding a ~0.5 μV acquisition amplitude resolution.

2.2. Stimulation

Three basic types of stimulation were used to assess the performance of the artifact removal framework: electrical micro-stimulation, electrical macro-stimulation and magnetic stimulation. All recording protocols were comprised of a period of spontaneous baseline activity followed by the application of stimulation.

2.2.1. Electrical micro-stimulation

Monopolar micro-stimulation generated using an optically isolated stimulator (STG-2008, Multichannel Systems, Reutlingen, Germany) was delivered using one of the microelectrodes (impedance 0.2–0.35 MΩ at 1 kHz). The stimulation pulses were current stabilized and consisted of 40–80 μA biphasic (200 μs cathodal followed by 200 μs anodal phase) pulses. The distance between the stimulating and recording electrodes varied and ranged from ~500 μm up to 5 mm.

Multiple patterns of stimulation were used for the assessment: (1) continuous stimulation, either at a high frequency (HF) (135 Hz) or low frequency (LF) (10 Hz), (2) bursts of stimulation—30 bursts of stimulation, separated by an interval of 500 ms; each burst was comprised of 40 pulses at 135 Hz, (3) continuous high frequency stimulation with variable intervals between pulses (5, 6, 7, 8, 9 ms, equally distributed, yielding an average frequency of 135 Hz).

2.2.2. Electrical macro-stimulation

Electrical macro-stimulation was used to simulate the pulses applied for therapeutic purposes during deep brain stimulation (DBS) in Parkinson's disease and other disorders. Monopolar macro-stimulation generated using an optically isolated stimulator (STG-2008, Multichannel Systems) was delivered through a Nary-lene coated 28 G tube (impedance 2 kΩ at 1 kHz). The stimulation pulses were voltage stabilized and consisted of 0.5–3 V monophasic (60 μs cathodal phase) or 2–6 V biphasic (200 μs cathodal followed by 200 μs anodal phase) pulses. These were applied at a continuous high frequency (125 Hz or 135 Hz) for 120 s. The distance between the stimulating and recording electrodes varied and ranged from ~500 μm up to 4 mm.

2.2.3. Magnetic stimulation

Magnetic stimulation pulses were generated using a custom-made magnetic stimulator. The stimulation was applied using a custom-made 25 mm diameter circular coil. The stimulation pulses were comprised of a cosine shape, with a 200 μs cycle length. The generated fields were up to 1.5 T using voltages of 200–800 V. The stimulation pulses were applied using the coil located within the recording chamber. Each stimulation session consisted of 60–120 pulses at a frequency of 0.2–4 Hz.

2.3. Simulated data

The simulated data used for validation of the artifact removal were comprised of recorded signals, to which simulated artifacts were added. The artifacts were based on the typical shape of electrical biphasic micro-stimulation artifacts of the form:

$$a(t) = e^{-\tau \cdot t} \cdot \mathrm{sinc}(\lambda \cdot t) \quad (1)$$

where τ is the time constant of the decaying exponent and λ is the cycle length of the sin c function. The mean values of τ and λ were manipulated, and the variability between consecutive artifacts was changed using different widths of the distributions of τ and λ. This function was chosen because of its resemblance to typical artifacts observed in real neuronal recorded data.

2.4. Software

The SARGE and its validation software were written using custom-made MATLAB (V2007B, Mathworks, Natick, MA) programs. The full framework is publicly available at: http://neurint.ls.biu.ac.il/software/SARGE. The SARGE has a graphical frontend in which all the stages of removal can be performed using a button click, following the adjustment of parameters within the GUI (Fig. 1). The software is extensible, allowing customization of the processing stages and import/export functions, if needed. The raw input files accepted by the SARGE are standard MATLAB data files, and there are no data acquisition requirements such as digitization rate or voltage range. No additional software other than MATLAB is required for the operation of the application.

3. Results

The framework is based on a multi-stage data flow to remove the stimulation artifacts. Briefly, the onsets of the stimulus pulses are detected and the signal is then divided into segments according to the detected pulses. A segment is defined as the part of the signal that starts at the pulse onset, and ends at the onset of the consecutive pulse. Later, the artifact at each segment is estimated and removed, and the artifact-free segments are concatenated to reconstruct the artifact-free signal. Finally, the artifact-free signal is filtered and an assessment of the quality of removal is conducted. At any given stage, the user can go back to execute previous stages of analysis, adjust parameters and improve the artifact removal quality. The process consists of sequential stages as follows (Fig. 2):

1. Pre-processing filter (optional).
2. Stimulus detection.
3. Artifact shape estimation.
4. Artifact removal.
5. Signal reconstruction.
6. Post-processing filter.
7. Quality assessment.

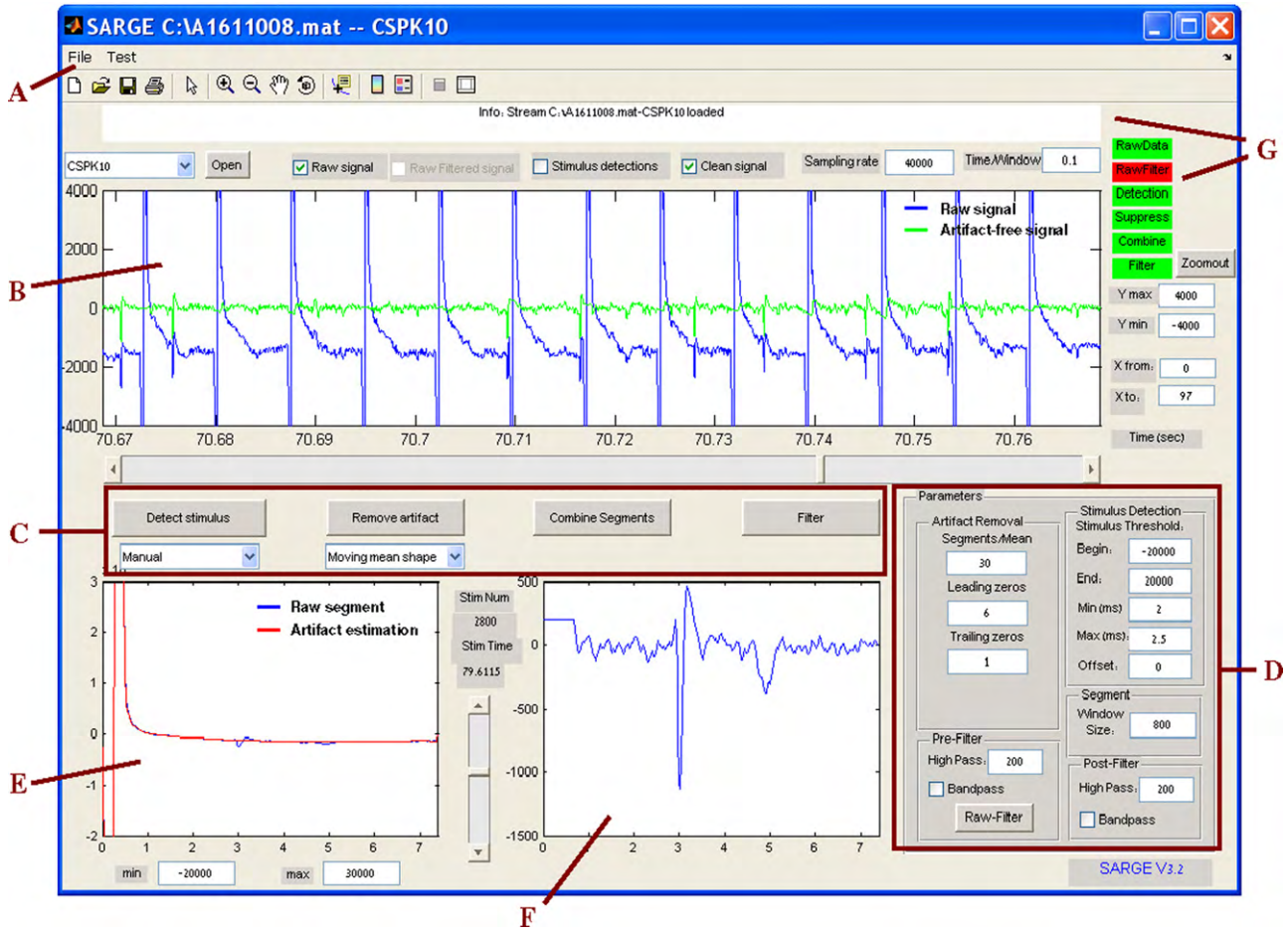

Fig. 1. The SARGE GUI. A screenshot of the SARGE application. (A) Menu options for loading data, exporting data, quality assessment, etc. (B) The main graphical view of the continuous data signals. The artifact-free signal is presented, superimposed on the raw signal. (C and D) Artifact removal stages (C): detection, estimation and removal, reconstruction and filtering. Each stage is performed using a button click, following the adjustment of parameters (D). (E) A graphical view of a raw single segment and its artifact estimation. (F) A graphical view of a single artifact-free segment. (G) Feedback buttons and text regarding the operations that were performed in the SARGE.

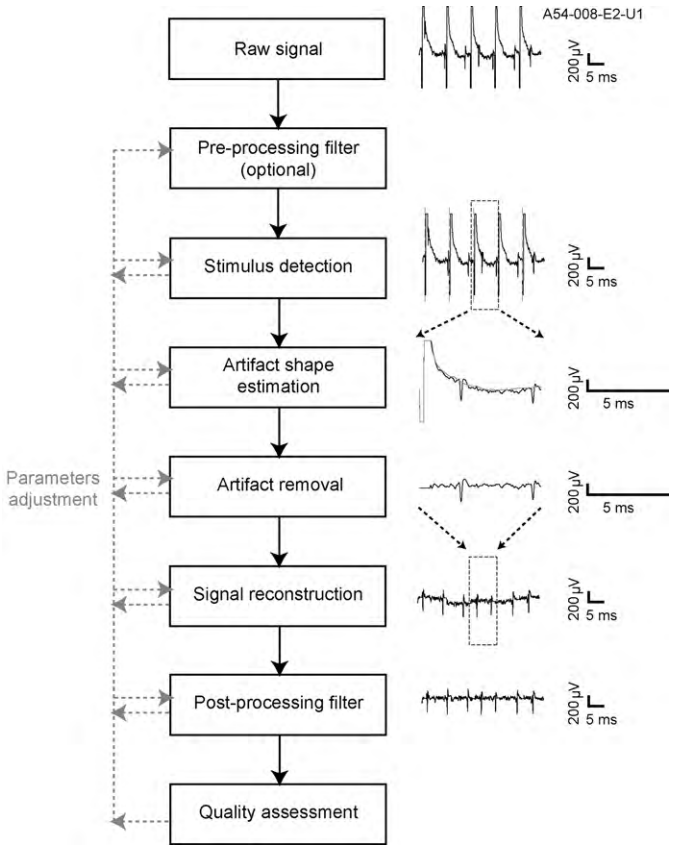

Fig. 2. Framework workflow scheme. A diagram of the analysis flow (left), accompanied by an example of a globus pallidus external segment (GPe) neuron recorded during local electrical high frequency micro-stimulation (right). Five stimulation pulses are presented in the raw signal, marked by gray dashed lines at the stimulus detection stage. The third pulse (framed) is enlarged, and the artifact shape (gray) is superimposed on the original signal (black) at the artifact shape estimation stage. The signal following the single pulse is presented after the artifact removal, and the segments of all pulses are then concatenated to reconstruct the signal. Finally, the signal is filtered to complete the artifact removal process, and then tested for its quality.

1. *Pre-processing filter*

The raw signal is dominated in some cases by low frequency changes, as a result of both the neuronal based activity (local field potential—LFP) and the slow components of the response to the stimulation pulses. These low frequency changes, when they significantly affect the raw signal, may impair the ability to properly remove the simulation artifacts. In such cases, it may be essential to filter the raw signal using a zero-phase high pass filter with a cutoff frequency well below the frequency typical to spiking activity. However, such a filter changes the original stimulus artifact shape and typically makes its temporal structure more complex, thus possibly making the next stages more problematic. Therefore, the pre-processing filter is used only when unavoidable.

2. *Stimulus detection*

Correct detection of the timing of stimulation pulse onset is crucial for subsequent removal. Clock synchronization of the stimulation and acquisition systems are the only way to guarantee exact timing between them. However, most acquisition systems work asynchronously with the stimulating system and are thus limited to synchronization at the timing level of the sampling interval (1/sampling rate). The simplest method for detection of the stimulation pulses is by using the digital pulses emitted by the stimulator marking the beginning and end of each stimulation pulse. However, in some cases these markers are unavailable, or their sampling rate does not match the sampling rate of the recorded signal, thus using them may lead to small but significant deviations from the exact timing of pulses. In these cases, the identification of the signal depends on identifying aspects of the structure of the artifact and typically the threshold crossing of the recorded signal is used for the detection of pulse onset.

Following the pulse onset, the recorded signal undergoes an initial period of considerable electrical changes, during which spiking activity cannot be detected at all. This *non-usable period* (NP) defines the part of the signal that will be omitted during the later stage of artifact estimation, as well as the remaining part of the segment (*usable period*, UP) in which the artifact will be estimated and spiking activity can be extracted following its removal.

The primary component of the NP, which starts immediately after pulse onset, is the saturation period (SP). However, instead of using the SP *per-se*, a voltage threshold crossing is used in the SARGE to determine the end of the NP for two reasons. First, the voltage changes at the first few samples (usually 1–3 samples for 40 kHz sampling rate) following the SP are typically very steep, and distort the estimation of the artifact. In these cases, the NP is always equal to or longer than the SP (NP = SP for voltage thresholds equal to the saturation values). Second, in some cases the signal does not reach saturation (e.g., when using electrical macro-stimulation and recording at a distant site). In the absence of the SP, voltage threshold must be used to determine the NP.

The voltage threshold can be adjusted manually within the SARGE GUI, and the threshold crossing can be limited to a predefined time range according to the pulse duration. The values of the thresholds (negative and/or positive) depend on the stimulus pulse shape (biphasic vs. monophasic, anodal vs. cathodal) and amplitude (Fig. 3). Additionally, they are affected by the distance between the stimulation and recording sites, in that weaker stimuli or larger distances generate smaller artifacts.

The framework enables the inspection of the quality of stimulus detection graphically, as one can view the raw signal with the identified stimulus pulses marked on it, as well as the enlarged and specific identification of each pulse separately.

3. *Artifact shape estimation*

Following the detection of the stimulus pulses, the signal is divided into post-stimulus segments. The signal in each segment (S) is presumably comprised of the sum of two components: the original (or neuronal) signal (X) and the stimulation artifact (A):

$$S_i = X_i + A_i \quad \forall i = 1 \ldots n \tag{2}$$

Non-additive interactions together with noise of the system also exist in the recorded signal. However, since they cannot be estimated appropriately, we use the simplified interaction that only includes the additive artifact. Since artifacts on simultaneously recorded channels may differ greatly, the artifacts are estimated for each channel separately and are not used across electrodes. The methods used to estimate the artifact (A_i) can be divided into two groups: averaging methods and function fitting methods.

3.1. *Averaging methods*

Averaging methods are based on the assumption that artifacts generated by different pulses are similar to each other over multiple segments, whereas the spiking activity within each segment is different. Therefore, the average signal over segments is a good estimator of the artifact shape regardless of its characteristics (shape, duration or magnitude).

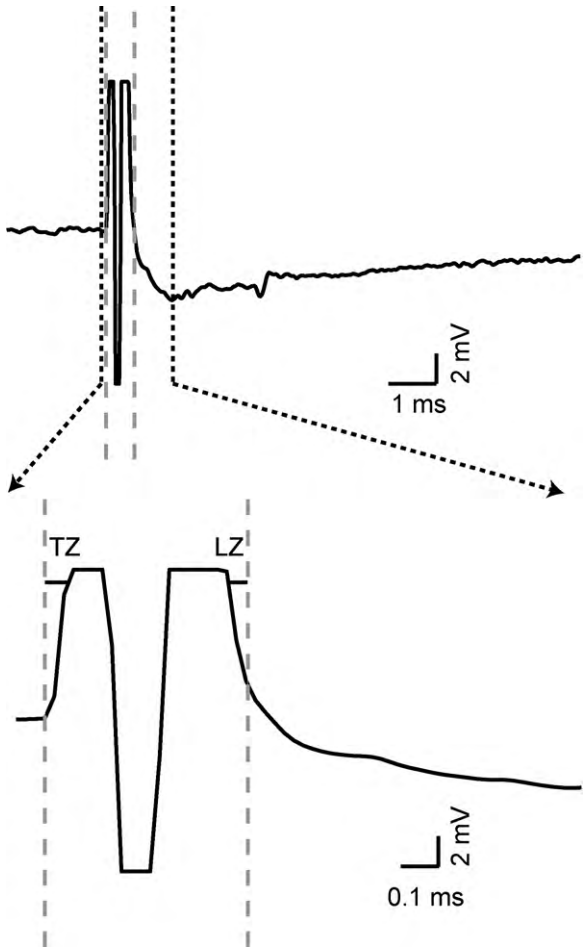

Fig. 3. Stimulus detection by threshold crossing. An example of a primary motor cortex (M1) signal recorded during magnetic stimulation. In the absence of digital markers for the onset of stimulus pulses, complex shapes of artifacts can be detected using thresholds and padded zeros for the pulse onset (left horizontal line, TZ—trailing zeros) and for the artifact ending (right horizontal line, LZ—leading zeros). Gray dashed lines indicate the resulting detection of pulse onset (left) and end of FNP (right).

3.1.1. Global average

The most basic averaging method is the estimation of the artifact ($\widetilde{A}$) by averaging the signal over all segments (Wichmann, 2000; Hashimoto et al., 2002; Montgomery, Jr. et al., 2005):

$$\widetilde{A} = < S > \quad (3)$$

This is the fastest method in terms of computation as $\widetilde{X}$ is calculated only once. However, it has several major limitations. The most prominent limitation is the high variability between artifact shapes over time in most signals. Thus, over long periods of time the average shape does not estimate the individual shapes well. Moreover, larger artifacts, and even longer NPs, are typically seen during the first tens of pulses of stimulation compared to the remaining pulses, mostly during high frequency stimulation. Since these artifacts are very large, they might affect the estimation of all artifacts, though they only occur in a small fraction of the segments.

3.1.1.1. Average for bursts of pulses. Stimulation protocols which are comprised of bursts of pulses (see for example electrical micro-stimulation protocol 2) result in major differences between consecutive segments in each burst and a great similarity between the segments which have the same index in all bursts; i.e., the *i*th segment in a burst is similar over all bursts. To overcome this inter-burst variability, averaging may be performed for each *i*th segment in a burst over all bursts (Bar-Gad et al., 2004).

3.1.2. Dynamic average

To overcome the dynamic changes in artifact characteristics, an averaged signal can be computed for small groups of consecutive segments. Segments which are temporally close to each other typically have similar stimulation artifact shapes; therefore their averaged signal can serve as a good estimator for the artifact. For each *i*th segment, the artifact is estimated as the averaged signal of $2k_w + 1$ segments (window size) at its vicinity (Fig. 4A–C):

$$\widetilde{A}_i = \frac{\sum_{j=i-k_w}^{i+k_w} S_j}{2k_w + 1} \quad (4)$$

where S_j is the *j*th segment. At the edges of the signal, where the number of segments before or after the *i*th segment are smaller than k_w, only the available segments are used and normalization is performed accordingly. The window size can be set manually within the SARGE GUI, and its value depends on multiple factors such as the frequency of stimulation, the total number of stimulation pulses, the variability of artifacts shapes during the stimulation train, and the required quality of the artifact-free signal. Typical values of window size are 30 for high frequency electrical stimulation of thousands of stimulation pulses (see example in Fig. 4B) and 5 for low frequency magnetic stimulation which has a total of tens to hundreds stimulation pulses.

The dynamic average method is slower than the global average method, as it requires multiple calculations of the averaged shape (the number of pulses). The quality of estimation is less affected by changes in the artifact shape over time; however it is still limited for two main reasons: (1) when the artifacts exhibit high variability, the window size needs to be reduced, and thus the calculated average is less reliable and (2) in some cases, spikes are evident after every stimulation pulse at a precise latency (~1 ms) relative to the stimulus onset. This perfect (100%) time-locking neuronal activity is most likely a result of direct axonal stimulation and antidromic propagation. In these cases, the shape of the spike becomes part of the average signal, and the extraction of this spiking activity becomes impossible, undermining the whole purpose of the artifact removal process. Thus, averaging methods cannot be used for the estimation of the artifacts. In other cases, where a non-precise time-locking firing is evident, probably as a result of synaptic transmission (≥2 ms after the pulse onset) (Erez et al., 2009), the probability of firing after each pulse is typically well below 100%, and there are jitters in the latency. The average pattern in these cases is mostly unaffected by the time-locking pattern of spiking activity allowing the usage of averaging methods.

Some variants of this method may be used, as described in the following subsections.

3.1.2.1. Dynamic average for varying lengths of segments. This variation is used when the inter-stimulus intervals are different in length (see for example electrical micro-stimulation protocol 3). In this case, only segments of the same length are considered for every calculated average; i.e., they are not necessarily consecutive. In order to use temporally close enough segments, the window size is typically reduced. The quality of artifact estimation declines due to both the smaller window size and the variability between non-consecutive segments.

3.1.2.2. Dynamic average for bursts of pulses. An improved averaging version for bursts of pulses (see Section 3.1.1.1) uses a dynamic average for each *i*th segment in a burst, where the average is comprised solely of $2k_w + 1$ bursts (window size), instead of all bursts (Fig. 4D and E).

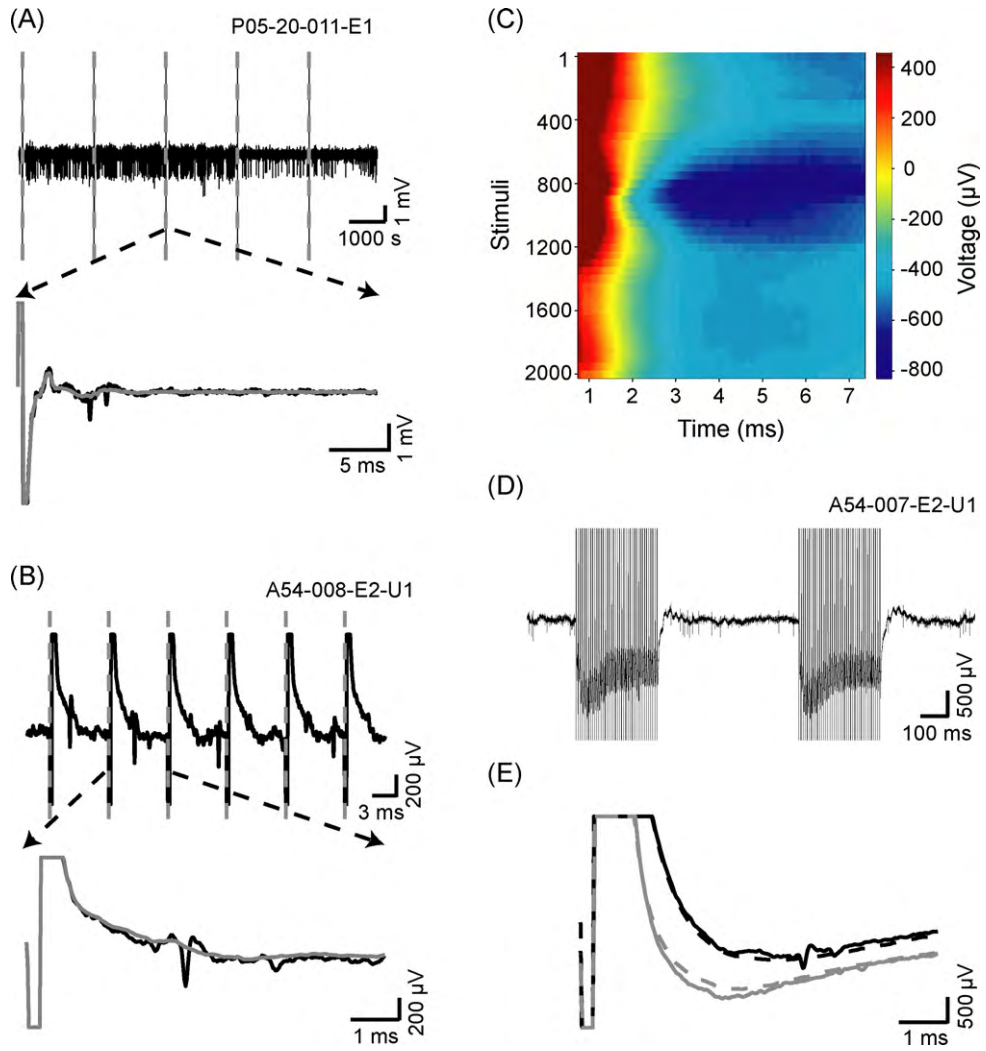

Fig. 4. Stimulation artifact estimation—averaging methods. (A and B) Similar shapes of consecutive artifacts (upper figures, pulses marked by gray dashed lines) may lead to a good artifact estimation of each single pulse (bottom, raw signal in black, artifact estimation in gray) using the dynamic average method. (A) An example of an M1 neuron recorded during magnetic stimulation (window size for estimation: 10). (B) An example of a GPe neuron recorded during local high frequency micro-stimulation (window size for estimation: 30). (C) An example of another GPe neuron recorded during local high frequency micro-stimulation, showing the mean estimated artifact over the stimulation period, as calculated by the dynamic average method. The FNP (0.75 ms) was excluded from the estimation. (D and E) Estimation by dynamic average for bursts of pulses. The same neuron from *B* during application of bursts of stimulation is presented. (D) Two bursts of high frequency, 40 pulses each, are shown (up), displaying similar inter-burst dynamics of stimulation artifacts. (E) Enlargements of the 1st and 11th post-pulse raw segments (black and gray, respectively) from the 1st burst from *D*, and the estimated artifacts (dashed black and gray lines, respectively) as calculated by the dynamic average for bursts of pulses method, demonstrating the inter-burst difference in artifact shapes. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of the article.)

3.2. Function fitting methods

Function fitting methods are based on a separate fitting of each stimulus pulse artifact to a function or a function family, such that the fit captures the artifact shape, but not the high frequency changes which characterize the spikes. Unlike averaging methods, the artifact estimation is not based on a common artifact template for a group of post-pulse segments, but rather on a unique fitting of a function for each segment. The fit is calculated on each post-pulse segment, following the exclusion of the NP. This estimation method is used primarily when the artifact shape changes rapidly over the stimulation period, or the number of pulses is not large enough for reliable averaging. In these cases, function fitting yields more accurate estimations; however its computation time is typically much longer because of the separate calculations performed for each post-pulse segment. These methods have another advantage due to their ability to remove the artifacts when the segments between pulses are not equal in length or when they are part of bursts of stimuli; namely, they do not require specific relations between the different segments. Additionally, these methods are effective for cases in which averaging methods are impractical, such as when very similar spiking activity is evident in most of the segments, thus included in the averaged signal and incorrectly identified as artifacts.

3.2.1. Exponential fit

The artifact shape is dominated by an exponential decay shape which is the outcome of the resistance–capacitance (RC) properties of multiple elements of the system. Therefore, the artifact shape may be theoretically described as the sum of exponential functions (Harding, 1991):

$$\widetilde{a}_i(t) = \sum_{j=0}^{n_e} c_{i,j} \cdot e^{-\lambda_{i,j} \cdot t} \tag{5}$$

where n_e is the number of exponential functions, and $c_{i,j}$ and $\lambda_{i,j}$ are the weight coefficient and rate parameter of the jth exponential function in the sum of fitted function to the ith segment.

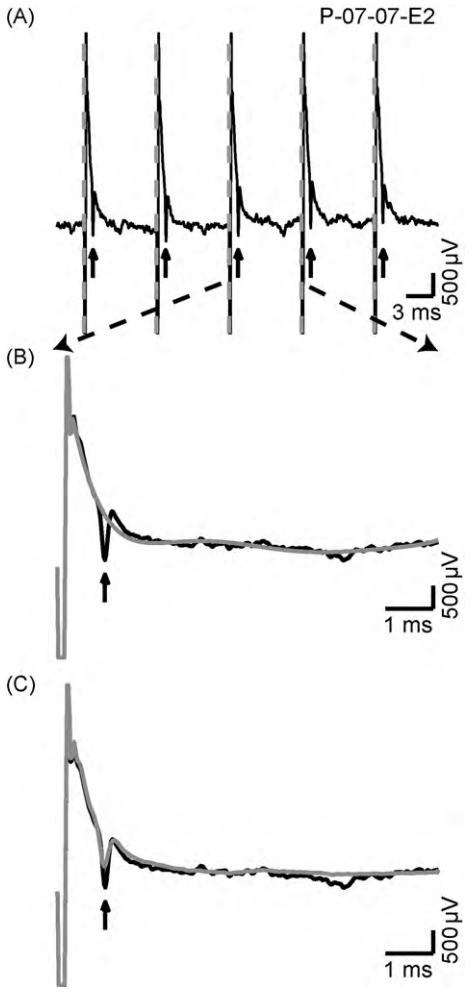

Fig. 5. Stimulation artifact estimation—function fitting. An example of a GPe neuron recorded during high frequency subthalamic nucleus (STN) macro-stimulation at 2 V, presenting a few consecutive stimuli (A) and a single artifact and its estimation using a 6th degree polynomial fit (B, black and gray, respectively). The time-locked response at a precise latency (spikes marked by arrows) following the stimulus pulse (marked by gray dashed lines) would be hidden when estimating the artifact using averaging (C).

The fit for each segment is minimized over the Euclidian distance between the signal and the estimated artifact:

$$\min\left[\sqrt{(S_i - \widetilde{A}_i)^2}\right] \tag{6}$$

3.2.2. Polynomial fit

A polynomial function can be fitted for each artifact of the form (Fig. 5):

$$\widetilde{a}_i(t) = \sum_{j=0}^{n_p} c_{i,j} \cdot t^j \tag{7}$$

where n_p is the degree of the polynomial functions, and $c_{i,j}$ are the coefficients of the polynomial function fitted to the ith segment. Like the exponential fit method, the fit for each segment is minimized over the Euclidian distance between the signal and the estimated artifact. Using this estimation, no preliminary assumptions regarding the artifact shape are required, and the degree of the polynomial can be adjusted as necessary. As the degree of the polynomial increases, the fit improves and the computation time increases. Two considerations should be taken into account when choosing the degree of the polynomial functions. First, over-fitting should be avoided, as the fit should only capture the simulation artifact, and not the spiking activity. Second, the degree of the polynomial primarily depends on the length of the segment, where a larger degree is required for longer segments. A polynomial degree of 8 usually yields good results for high frequency stimulation, where a larger degree (11–13) is required for low frequency stimulation.

A variant of this method may be used when removing artifacts for low frequency stimulations, in which the segments between pulses are long and the artifacts comprise only a small fraction of the segment. A polynomial function is then fitted to the beginning of the segment, rather than to the whole segment. The length of the segment for which the polynomial is fitted can be adjusted (usually 10–20 ms), and the rest of the signal in each segment is later concatenated as is.

3.2.3. Local polynomial fit

A polynomial function may be dynamically fitted for sub-segments in each post-pulse segment (Wagenaar and Potter, 2002). The fitting process is as follows: for each sampled point (t_c) in the ith segment (S_i), a $2w+1$ sample sub-segment ($S_i^{t_c}$) around it is defined as follows:

$$S_i^{t_c} = [s_i(t_c - w), \ldots, s_i(t_c + w)] \tag{8}$$

A polynomial function is then fitted for this sub-segment:

$$\widetilde{a}_i^{t_c}(t) = \sum_{j=0}^{n_p} c_{i,j}^{t_c} \cdot (t - t_c)^j \tag{9}$$

where n_p is the degree of the polynomial functions, and $c_{i,j}^{t_c}$ are the coefficients of the polynomial function fitted to the sub-segment. The fit for each sub-segment is minimized over the Euclidian distance between the sub-segment signal and the estimated sub-segment artifact. Finally, the value at each point t_c is estimated based on this fit:

$$\widetilde{a}_i(t) = \widetilde{a}_i^{t_c}(t) \text{ for } t = t_c \tag{10}$$

At the edges of each segment, in which there are not enough samples before or after t_c, a polynomial fit for a sub-segment of length $w+1$ only is used.

3.3. Artifact estimation at the segments' edges

The signal values at the beginning and end of each segment are in some cases characterized by the large rapid electrical transients that may display as new artifacts in the processed signal. Different factors may lead to these transients, such as jitters of triggering points of the stimulation onsets. These new artifacts may be mistakenly identified as spikes during the sorting process following the artifact removal. The omission of the first few samples of the segment following the NP, termed 'Leading zeros' (LZ) and the last few samples of the segment, termed 'Trailing zeros' (TZ), usually leads to an improved estimation and may prevent the generation of these new artifacts. For each segment, the sum of NP, LZ and TZ defines the *full NP* (FNP), which is the part of the segment that is omitted during the estimation of the artifacts. Longer FNP may improve the quality of artifact removal, but spiking activity that might occur during that period cannot be extracted at the consequent offline spike sorting. Moreover, even spikes that only part of their action potential occurred during the FNP might be misdetected during the spike sorting process, thus resulting in an actual longer FNP, depending on the action potential length.

The LZ and TZ parameters can be adjusted during the estimation process by setting their values manually within the SARGE GUI. Typical values for LZ are 4–8 sample points (i.e., 0.1–0.2 ms at a 40 kHz sampling rate) for high frequency electrical stimulation, but

in some cases (such as magnetic stimulation) values larger than 10 may be used. The TZ value is typically set to 1 (0.025 ms at a 40 kHz sampling rate), but values of 2–3 may be used in some cases. For magnetic stimulation, the typically used TZ value is 5 (0.125 ms at a 40 kHz sampling rate), and may be larger than 10 in some cases.

The graphical tools of the SARGE enable the user to view the result artifact-free signal and assess its quality and the generation of new artifacts. However, in some cases the generation of these artifacts cannot be prevented completely, and their identification as spikes mostly depends on the consequent offline spike sorting process.

4. *Stimulation artifact removal*

Given the estimation of each stimulus artifact and regardless of the estimation method, the artifact in each segment is removed by subtraction:

$$\tilde{X}_i = S_i - \tilde{A}_i \quad \forall i = 1 \ldots n \tag{11}$$

where S_i is the original signal of the ith segment, and $\tilde{X}_i$ is the artifact-free signal of the ith segment. Following the subtraction, the artifact-free signal is composed of separate segments of the estimated neurophysiological activity (Fig. 6A–C).

The SARGE framework provides graphical tools for viewing the segmented original and artifact-free signals. At this stage, the quality of the artifact-free signals is assessed at the segment level by visual inspection by the experimenter or by various statistical measures (see below for details in the quality assessment section).

5. *Signal reconstruction*

The next stage in the process is reconstruction of the signal by concatenating the artifact-free segments into a continuous signal. Two main problems must be dealt with at this stage. First, the segments following artifact removal are shorter than the original segments that were separated by the stimulus pulse onset since they do not include the FNP. Second, the signal values at the end of one segment and at the beginning of the next segment are not necessarily (and actually, in most cases are not) the same. A simple solution for these problems is a naïve point-to-point linear interpolation, from the last point in the nth segment to the first point in the $n+1$th segment (Fig. 6D and E). It is important to note that this may alter multiple properties of the signal such as its spectral content; thus, more advanced solutions are required in some cases.

6. *Post-processing filter*

The concatenation of the segments inserts artificial components because the intervals between the segments at the FNP are filled to preserve the timing of the spiking activity as in the original signal. Therefore, the signal following the concatenation may include low frequency elements (Fig. 6E and F). A zero-phase high pass filter is used to remove these low frequency artificial changes. The filter inevitably changes the signal and the embedded spike shapes within it. However, since such a filter is typically used before offline spike sorting, it is not an exceptional alteration within this context. Filtering is the last stage in the artifact removal process, after which the signal is subjected to final quality assessment.

7. *Quality assessment*

The quality of the resulting artifact-free segments and continuous signal is assessed using a battery of tools supplied by the framework, for both subjective and objective assessment. Following the assessment, the parameters of any of the stages in the process may be adjusted. The whole artifact removal process, or parts of it, is then performed again, until acceptable results are achieved.

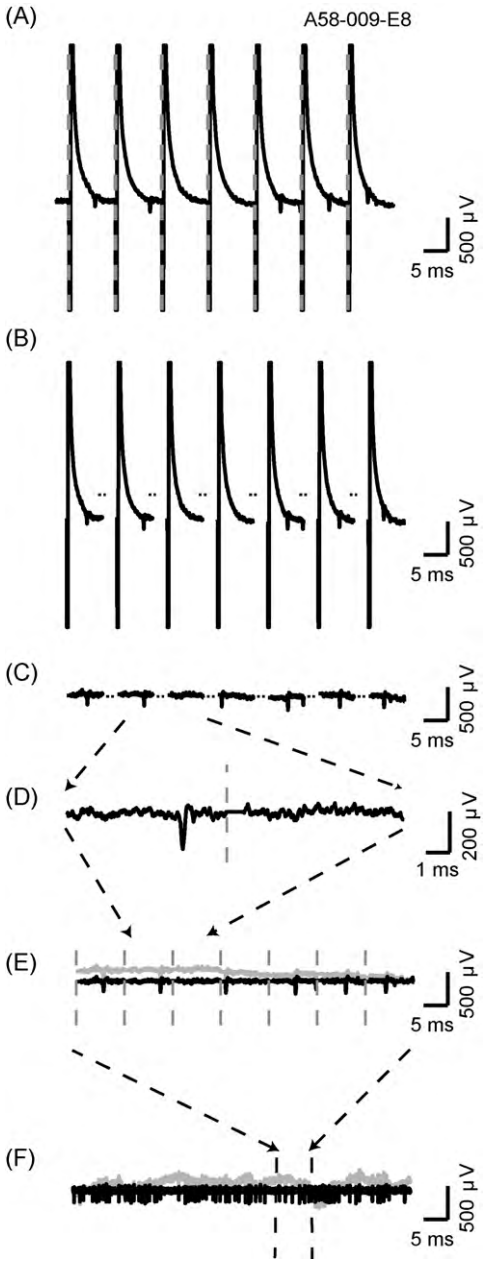

Fig. 6. Stimulation artifact removal, signal reconstruction and filtering. (A) Raw signal (black) and stimulation pulse detection (dashed gray) of seven consecutive pulses, recorded from a GPe neuron during local high frequency micro-stimulation. (B) Division of the signal into segments. (C) The artifact-free segments. (D) An enlargement of the 2nd and 3rd segments, showing the linear interpolation during the reconstruction of the signal. Stimulus pulse detection is indicated by dashed gray line. (E) The reconstructed signal (gray) and the same signal following filtering (black) using a 200 Hz four-pole Butterworth filter. (F) Zoomed out view of the same signal, demonstrating the need for filtering. Throughout the figure: gray dashed lines indicate pulses of stimulation.

7.1 *Subjective assessment of artifact removal*

The resulting signal is viewed, and can be superimposed on the original signal and/or marked with the stimulus pulse onset. The aim of this subjective view is to assess the ability to extract spiking activity out of the signal, with minimal interference from stimulus

artifact residuals. Problematic suppression of artifacts is in many cases reflected as large transients in the artifact-free signal at the frequency of stimulation (Fig. 7A). Usually, such transients appear in parts of the signal where the artifacts were not stable over time, or where relatively large artifacts occurred. In these cases, the individual segments in these parts are viewed, and the artifact removal process is corrected by adjusting parameters (Fig. 7B).

Other tools provide the experimenter with the ability to assess the neuronal response to stimulation. If the recorded neuron is large enough relative to the background activity, a threshold can be used to roughly identify its spikes, and a peri-stimulus-time histogram (PSTH) can be generated, showing whether the neuron tends to fire at a specific latency following the stimulus pulse (time-locked response). In other cases, all or part of the artifact-free segments can be presented, aligned to the stimulus onset (Fig. 7C). Using this presentation, a time-locked response can be observed, as the spikes are concentrated around a specific latency relative to the pulse onset. These tools provide a preliminary assessment of the neuronal activity.

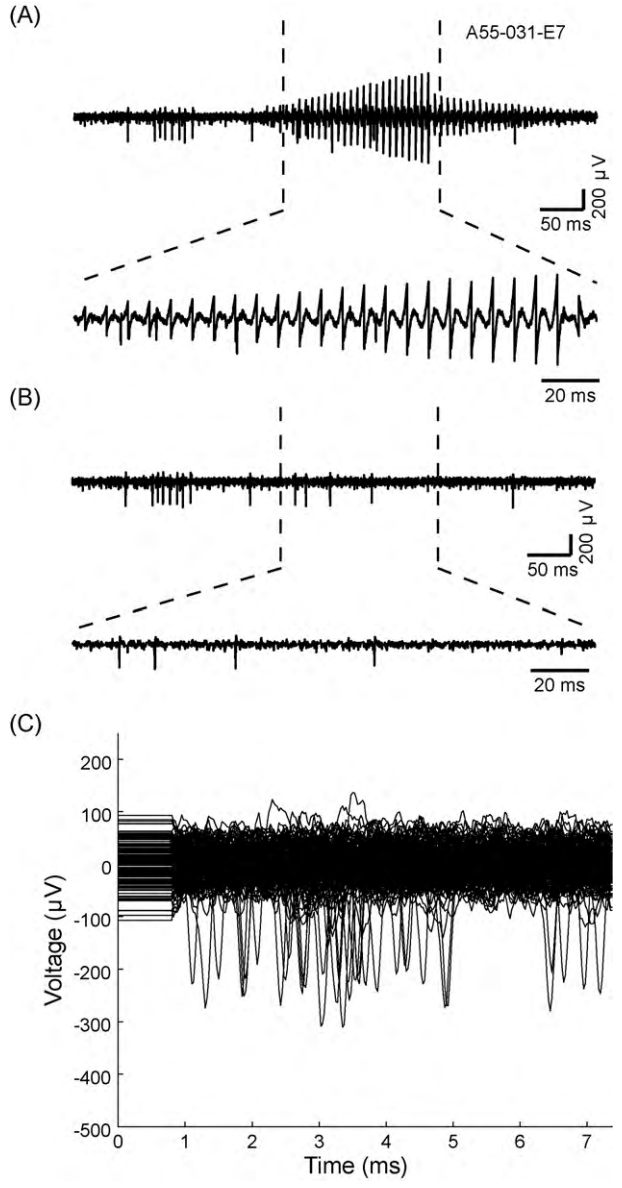

Fig. 7. Quality assessment. (A and B) An artifact-free signal recorded from a GPe neuron during local high frequency micro-stimulation, using 4 samples of LZ (A) which results in less than optimal artifact removal, and using 12 samples of LZ (B) which drastically improves the removal of the artifacts. The artifacts were removed using the dynamic average method; the signal at the bottom is a zoomed-in view of the upper signal. (C) Segmented artifact-free signal of the same neuron aligned to the stimulus onset. Segments of 150 consecutive pulses are superimposed, demonstrating the time-locked response to stimulation.

7.2 *Objective criteria for artifact removal*

To quantify the quality of the artifact removal process for a given signal, we have developed several measures.

7.2.1 *Segments' FNP distribution*—During the FNP, spiking activity cannot be identified. The FNP varies between different types of stimulations, and is affected by a variety of parameters, such as the pulse duration and amplitude, the capacitance of the acquisition device, the distance between the stimulation and recording sites, etc. Generally, the optimal situation is to have FNPs as short as possible. However, increasing the number of leading zeros, which directly leads to the lengthening of the FNP, is commonly used to overcome problematic artifacts that cannot be estimated properly. Therefore, in most cases there is a trade-off between a short FNP and a satisfactory removal of the artifacts. The distribution of the FNPs over the segments of the signal can be used as part of a battery of quality measurements, where smaller mean and median values of the distribution indicate better performance and skewed distributions indicate problematic artifact suppression of a subset of the segments (Fig. 8A and D).

7.2.2 *Distribution of root mean square (RMS) of signal segments*—RMS can be calculated for each segment on the difference between the raw segment and the estimated artifact, after exclusion of the FNP:

$$\mathrm{RMS}_i = \sqrt{\frac{(S_i - \mathring{A}_i)^2}{l_i}} \quad (12)$$

where l_i is the number of samples in the ith segment.

The distribution of the RMS values of the signal segments is a measure of quality of the artifact estimation, where smaller mean and median values indicate better artifact removal and skewed distributions indicate problematic artifact suppression of a subset of the segments (Fig. 8B and E).

7.2.3 *Distribution of peak-to-trough (PTT) difference*—The values of the first few samples of each artifact-free segment following the FNP are usually the most indicative of the quality of estimation, as the largest deviations from the artifact occur at these points. The peak-to-trough difference of the ith segment is defined as follows:

$$\Delta\mathrm{PTT}_i = \mathrm{abs}[\max[x_i(1), \ \ldots, x_i(m)] - \min[x_i(1), \ \ldots, x_i(m)]] \quad (13)$$

where $x_i(1), \ldots, x_i(\mathrm{m})$ are the first m (window size) samples in the ith artifact-free segment following the FNP. Smaller values indicate better estimation, as can be assessed using the ΔPTT distribution and its mean and median (Fig. 8C and F).

Export data and metadata

Following the completion of the artifact removal process, the data and metadata can be exported for further analysis and use. The data include the artifact-free signal, the stimulation pulses onsets and the FNP of all segments. The metadata include all the parameters used within the SARGE application during the removal process.

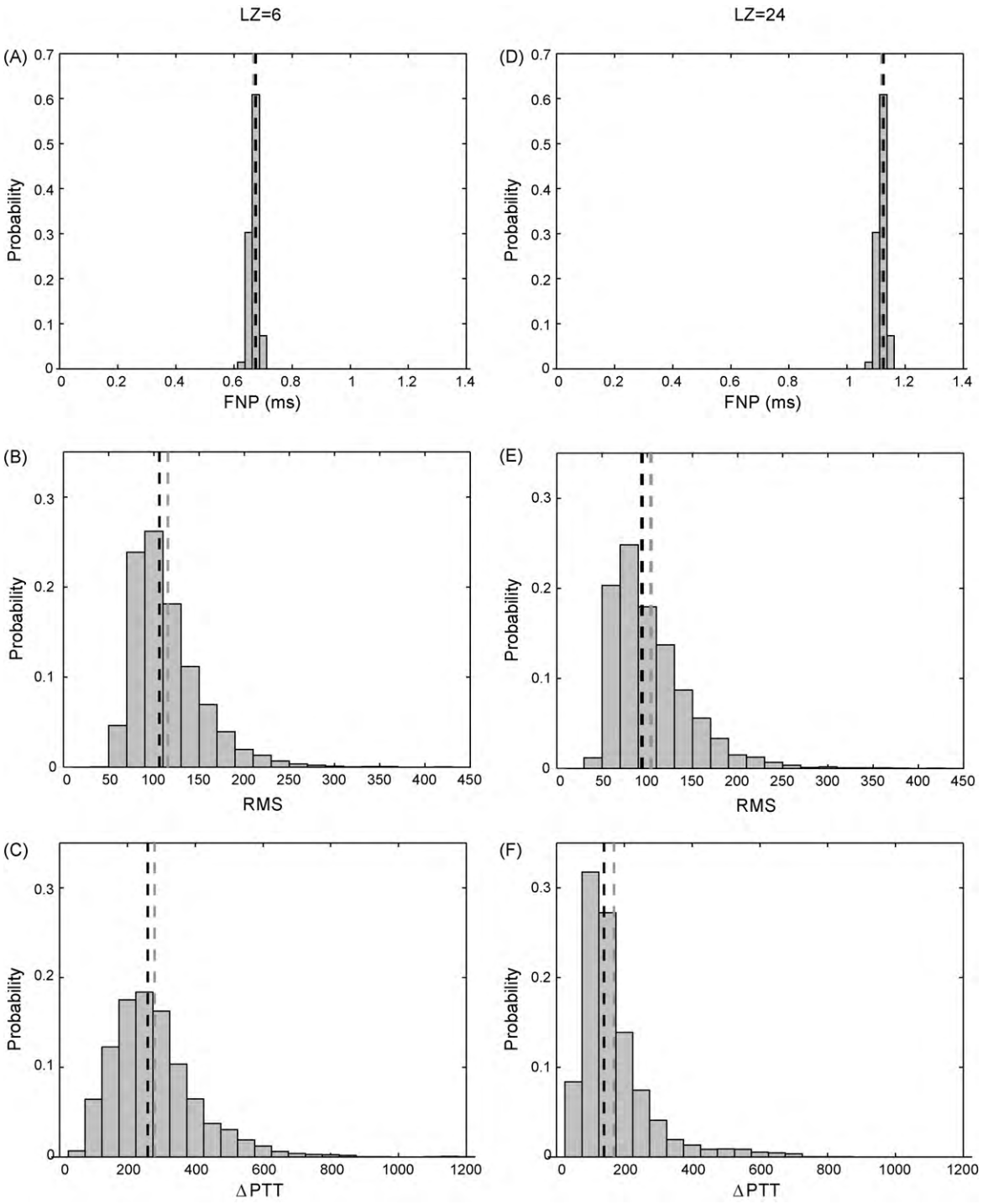

Fig. 8. Objective criteria for artifact removal. Segments' FNP (A and D), RMS (B and E) and ΔPTT (C and F) distributions of a GPe neuron recorded during local high frequency (135 Hz) micro-stimulation, for artifact removal using dynamic average with 6 and 24 LZ, respectively. Throughout the figure: black and gray dashed lines indicate mean and median of distributions, respectively. The removal of artifacts using 6 LZ yields unsatisfying results, where the use of 24 LZ improves the process results. The increase in the number of LZ is reflected in the larger values of FNP, and in the smaller values of RMS and ΔPTT. The narrow distributions of FNP indicate similar FNPs over the session's stimulus pulses.

The SARGE software is extensible, thus the exported data can be customized.

Validation of artifact removal quality

To validate the quality of artifact removal techniques presented in this study, we conducted three assessment-studies. Unlike the quality assessment tools, the validation studies are not part of the framework itself.

1. *FNP distributions of experimental data*—We calculated the FNP distributions of populations of recorded neurons during different types of stimulation used in this study: electrical micro- and macro-stimulation and magnetic stimulation (Fig. 9). All neurons went through the process of stimulation artifact removal and the resulting artifact-free signals were assessed by the experimenter as high-quality and passed the criteria in terms of the ability to correctly identify spiking activity. The FNPs over the population

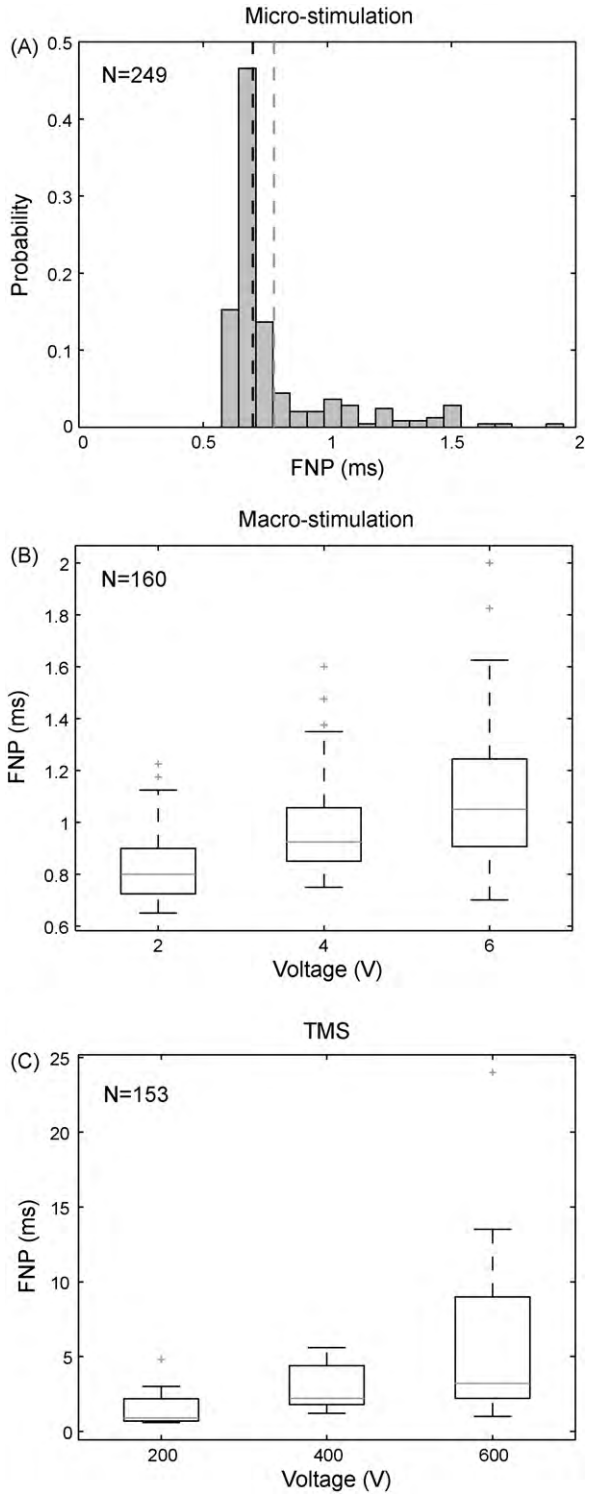

Fig. 9. Assessment of artifact removal over population. (A) FNP distribution for a population of pallidal cells recorded during pallidal high frequency (135 Hz) electrical micro-stimulation. Black and gray dashed lines indicate mean and median of distribution, respectively. (B) Box plot of FNPs distributions for a population of pallidal neurons recorded during pallidal high frequency (135 Hz) electrical macro-stimulation. (C) Box plot of FNPs distributions for a population of M1 neurons recorded during magnetic stimulation at 0.5 Hz.

provide an indication of the accuracy of the estimation methods used in the framework.

2. *Experimental data assessment*—We used a recorded signal with background unit activity (Moran and Bar-Gad, 2010) and no identifiable single units, and characterized its statistical properties before and during stimulation. We used this type of signal in order to avoid a distortion of signal characteristics as a result of the effect of stimulation on spiking activity. Optimal removal of artifacts should result in similar characteristics of the signal during stimulation compared to its pre-stimulation baseline properties. We used two measures to characterize the signal based on the distribution of the signal values and the spectral properties of the signal.

a. Distribution of signal values. Two distributions were constructed, before and during stimulation. Improper removal of artifacts resulted in high and low signal values seen as long tails of the distribution during stimulation, compared to the distribution of signal values before stimulation. To assess these tails, we calculated lower and upper significance bounds on the distribution of values before stimulation ($\alpha = 0.01\%$). The percentages of values in the distribution during stimulation that crossed these significance bounds were then calculated, where smaller percentages represent shorter tails of the distribution, implying better artifact removal.

b. Spectral analysis. Improper removal of artifacts yields large transients of signal values in the vicinity of each pulse, resulting in enhanced power at the frequency of stimulation. Two power spectral density (PSD) functions of the recorded signal were calculated, before and during stimulation. The total power of the PSD functions, before and during stimulation, in a 10 Hz window size around the frequency of stimulation was calculated, and the ratio between them was used to assess the artifact removal. This measure requires elimination of power grid activity (50/60 Hz); therefore it is not suitable for stimulation frequencies around this base frequency.

These measures were calculated for signals that were generated using different artifact estimation methods and/or parameters, and were then used to validate the artifact removal process as assessed subjectively by a human expert experimenter. Fig. 10 shows an example of an artifact-free signal processed using a varying number of LZ. Large transients are evident in the signal traces when using small values of LZ, and these transients disappear when increasing the number of LZ, indicating better artifact removal as could be assessed by the experimenter (Fig. 10A). This subjective assessment is reflected in the measures of the distribution of the signal values (Fig. 10B) and in the spectral analysis (Fig. 10C and D), thus validating the quality of the artifact removal process. The improved artifact removal as a result of the increased number of LZ is traded-off with a lengthening of the FNP at the expense of the UP.

3. *Combined experimental and simulated data assessment*—Artificially generated artifacts were arbitrarily inserted into a signal of a neuron that was recorded off-stimulation (Fig. 11A, see Materials and Methods, 'Simulated data' section, for details). The artificial artifacts were based on estimated artifacts recorded during electrical micro-stimulation. The time constant of the exponential decaying part (τ) and the sine function cycle parameter (λ) of the artifact were manipulated, thus creating different levels of variability between artifacts (Fig. 11B). The variable artifacts appeared either in a random order or sorted along the stimulation period to simulate the slowly changing artifact shapes over time, which are typically evident in recorded data. The artifacts were consequently removed using different methods and parameters, and the resulting signals were compared to the original signal

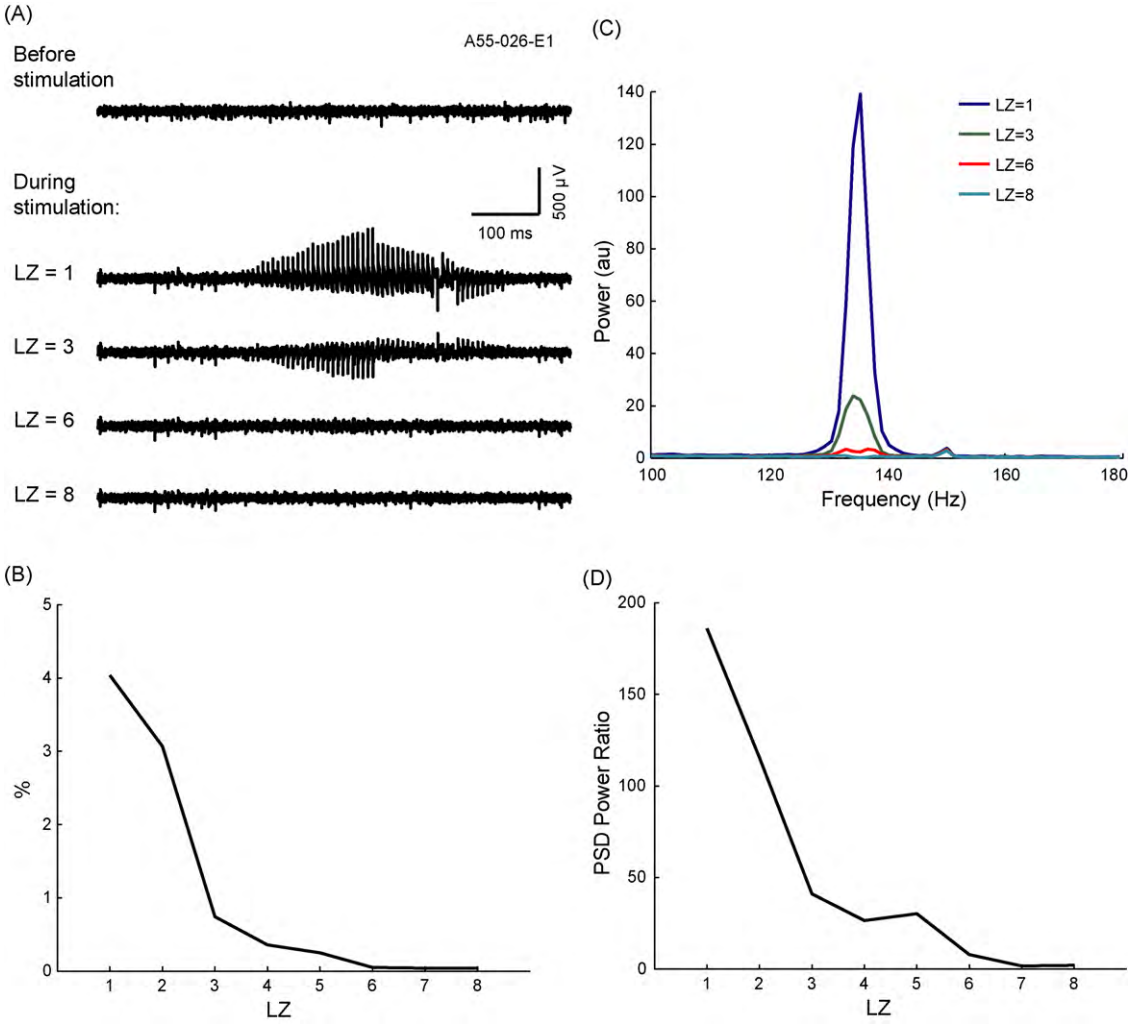

Fig. 10. Validation using a spike-free signal. Artifacts were removed from a signal recorded from the globus pallidus internal segment (GPi) during GPe high frequency (135 Hz) electrical micro-stimulation with no spiking activity to validate the artifact removal process. (A) Examples of the signal before and during stimulation, following the removal of the artifacts using the dynamic average fitting method and a four-pole Butterworth high pass (60 Hz) filter are presented. The number of LZ was manipulated to achieve a higher quality of removal. Using 1 or 3 LZ yielded high frequency large transients at the frequency of stimulation, which were abolished when using 6 LZ. (B) The percentages of the signal value distribution during stimulation that crossed the threshold significance upper and lower bounds ($\alpha = 0.01\%$) of the value distribution before stimulation, as a function of the number of LZ. (C) Power spectral density functions during stimulation. (D) The ratio between the power at the 130–140 Hz band during and before stimulation as a function of the number of LZ. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of the article.)

before the artificial insertion of stimulation artifacts. The same measures (distribution of values and spectral analysis) were used for this comparison (Fig. 11C and D). Artifacts with low variability were removed successfully by the three estimation methods, as reflected in the low percentages of significance-bound crossings and low values of the PSD power ratio around the stimulation frequency. However, increasing the variability between the artifacts resulted in the failure of the 'Global average' method and yielded different results for the two other methods tested. When the artifacts varied slowly over time (when temporally close artifacts had similar shapes), the 'Dynamic mean' method, which captures the average artifacts over small consecutive groups, was able to successfully remove the artifacts. This method however failed to properly remove the artifacts when the same artifacts were randomized with low similarity between consecutive artifacts. The 'Polynomial fit' method, which fits a function to each segment separately, does not depend on the variability between artifacts or the similarity between consecutive ones, and could successfully remove the artifacts in all data conditions. The estimation methods differed from each other not only in their ability to successfully remove different types of artifacts, but in their computation time as well (Fig. 11E).

4. Discussion

In this manuscript, we presented SARGE, a generalized framework for the removal of artifacts resulting from electrical and magnetic stimulation. The framework provides a generic infrastructure for a multi-stage semi-automatic artifact removal process, starting from a raw signal, through the detection of the stimulus pulses, the estimation of the artifacts, their removal and the reconstruction of the artifact-free signal. The framework provides a wide range of methods for the estimation of the artifacts and the adjustment of relevant parameters, as well as graphical tools and objective measures for quality assessment of the result signal.

This framework is an efficient and effective environment for the artifact removal process. On one hand, stimulation artifacts share similar characteristics; therefore the same building blocks and processing stages can be used for all of them. On the other hand, the artifacts of each signal, even from simultaneously recorded channels, are different from each other, and therefore a semi-automatic

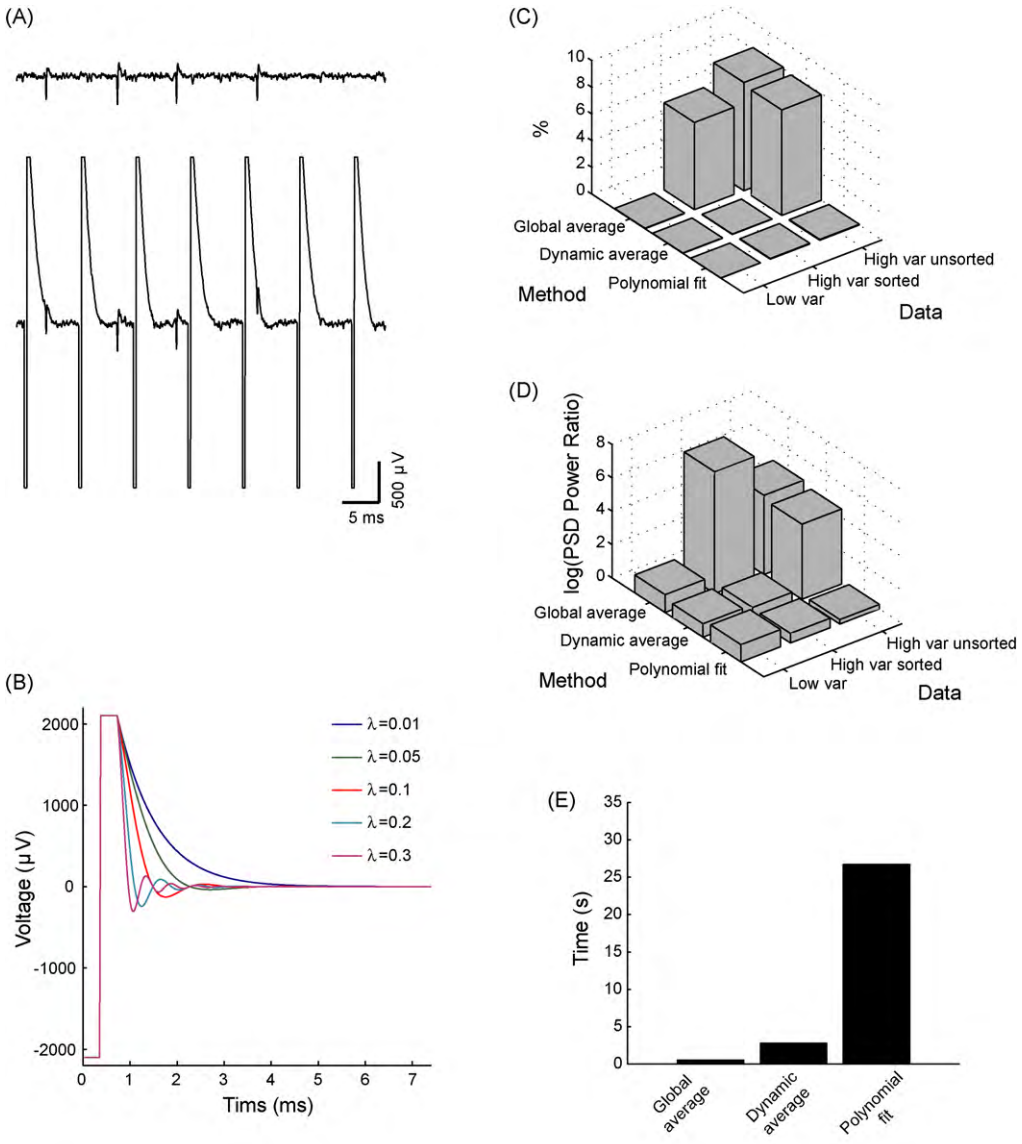

Fig. 11. Validation using a combined experimental-simulated signal. Artificially generated artifacts were inserted into a recorded signal and were then removed using three different methods: 'Global average', 'Dynamic average' and 'Polynomial fit'. The degree of variability between artifacts was manipulated, as well as the similarity of shapes between consecutive artifacts ('Low var', 'High var sorted'—slowly changing artifacts, and 'High var unsorted'—randomized artifacts). (A) An example of a raw trace from a recorded signal before (top) and after (bottom) the insertion of the artifacts. (B) Examples of the artifacts that were inserted into the signal, aimed at mimicking biphasic artifacts generated by high frequency electrical micro-stimulation. (C) Percentages of post-artifact removal signal value distribution that crossed the upper and lower threshold significance bounds ($\alpha = 0.01\%$) of the raw signal value distribution. (D) The ratio between the PSD power values at the 130–140 Hz band during stimulation compared to the raw signal. (E) Computation time required for the different stimulation artifact estimation methods, as calculated on the simulated data. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of the article.)

process rather than a fully automatic one is essential. Moreover, it can cope with varying degrees of difficulty in artifact removal which mainly depend on the shape and amplitude of the artifacts. These are affected by the distance between the recording and stimulation electrodes, the impedance and shape of electrodes, etc.

Some of the methods for artifact estimation used in this study are based on existing methods (Harding, 1991; Wichmann, 2000; Wagenaar and Potter, 2002; Hashimoto et al., 2002; Montgomery, Jr. et al., 2005) while others are novel. All the methods were generalized to handle different stimulation protocols and customized to common scenarios. The framework generates a generic workflow while combining automatic "objective" algorithms with human based "subjective" observation and settings. Moreover, this framework constitutes an innovative concept in regards to the artifact removal process, and marks a shift from "algorithm" and "data" centric approach to a "workflow" centric approach. The process is generalized at multiple levels. First, the framework is used for the removal of artifacts resulting from different types of stimulation such as electrical micro- and macro-stimulation and magnetic stimulation. Second, the framework enables the use of different artifact estimation methods for the same signal and the comparable inspection of the resulting artifact-free signal. Third, the experimenter is provided with tools for quality assessment of the resulting signal, including the option to adjust parameters to obtain better results. Finally, the framework provides an encapsulated workspace for the multi-stage process of artifact removal in which the experimenter can easily and conveniently start the process, go through all its stages, and finally perform quality assessment on the resulting signal.

The wide range of estimation methods makes it possible to tailor the most suitable method for a given signal based on the artifact shapes and the stimulation protocol pattern. The estimation

Table 1
Performance of artifact shape estimation methods.

Estimation method	Computation time	Changes in artifacts shape throughout stimulation	Suitability for bursts of stimulation	Suitability for non-fixed inter-stimulus intervals
Averaging methods				
Global average	++	−−	−	−
Average for bursts of pulses	++	−−	+	−
Dynamic average	+	−	−	−
Dynamic average for varying lengths of segments	+	−	−	+
Dynamic average for bursts of pulses	+	−	+	−
Function fitting methods				
Exponential fit	−	+	+	+
Polynomial fit	−	+	+	+
Local polynomial fit	−−	+	+	+

Computation time: ++ very fast, + fast, − slow, −− very slow.
Changes in artifacts shape throughout stimulation: + rapid dynamic changes, − slow dynamic changes, −− no changes.
Suitability for bursts of stimulation: + yes and − no.
Suitability for non-fixed inter-stimulus intervals: + yes and − no.

method and parameters which are used for a recorded signal also depend on the familiarity of the experimenter with the specific data and similar data sets. The accuracy of artifact shape estimation can be enhanced, but at a cost of greater manual time for parameter setting or computation time. These considerations, summarized in Table 1 and exemplified in Fig. 11, should be taken into account when choosing a method of estimation. Averaging methods, which are the fastest, are usually chosen when the artifacts are similar to each other. Function fitting methods, which fit a different function for each segment, require longer computation time, and usually yield better results when the artifacts change rapidly during the stimulation period. However, for complex shapes of artifacts which change dynamically, these methods might be the only ones that can yield accurate enough artifact estimation. Overall, methods with a relatively good accuracy/time ratio, such as dynamic averaging and polynomial fit, are the most commonly used. The total time it takes to remove artifacts from a record depends on two main factors. The first is the actual computation time, limited by the computer hardware. The second factor is the operations performed by the human experimenter, including data observation and parameters setting. Other factors that affect the total time of artifact removal include the duration of the recording, the number of stimulation pulses, the complexity of artifacts, the estimation method which was used, the required quality of artifact-free signal, etc. Roughly, for a skilled experimenter, it may range from less than a minute for a signal with standard artifacts up to over ten minutes in some cases of non-standard artifacts.

The framework provides a set of graphical tools for the assessment of the quality of the artifact-free signal, as well as objective comparable quantitative measures. The objective measures provide a good estimation of a successful process of artifact removal. However, there is no clear-cut answer to the key question what constitutes good removal. In other words, assessment measures can be used to compare methods operating on the same signal, but do not generalize well to provide normalized and absolute values. Despite considerable efforts, none of the objective measures was found appropriate to assess the goodness of the artifact-free signal in terms of its ability to extract the spiking activity during subsequent offline sorting procedures. Therefore, though an automatic process for artifact removal from multiple signals is the fastest way to handle the stimulation artifacts problem, especially for large data sets, the subjective assessment of a human observer cannot be replaced and is a necessary for complementary verification of successful artifact removal.

The concept of a framework-based artifact removal is novel and stems from the widespread use of stimulation in neurophysiology studies, including different types of stimulation types and patterns. This has raised the need for a generic solution to the artifact removal problem that can replace the specific solutions that have been suggested for special cases. This solution is part of a more general trend of neurophysiology research, in which transitions from single-case to integrated-based solutions are introduced to deal with the rapidly increasing types and amounts of data and software applications. For example, many spike offline sorting algorithms were originally introduced to handle different types of data, and later have been encapsulated in a generic graphical user interface that can support a wide range and variety of types of data (e.g., Plexon OFS). Other examples include NeuroML, which is used for the representation of neurons for modeling purposes and NueroShare, which is used for handling file formats from different software programs. From this perspective, the framework presented in this study is a step towards a generic solution for the stimulation artifact removal problem, and parallels the growth of stimulation-based studies in the field of neurophysiology research.

Acknowledgements

This study was supported in part by Israel Science Foundation (ISF) grant (1000-05), ISF Converging Technologies grant (1698-07) and Ministry of Health (MOH) grant (3-4033).

References

Bar-Gad I, Elias S, Vaadia E, Bergman H. Complex locking rather than complete cessation of neuronal activity in the globus pallidus of a 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine-treated primate in response to pallidal microstimulation. J Neurosci 2004;24:9410–9.

Benabid AL, Pollak P, Gross C, Hoffmann D, Benazzouz A, Gao DM, et al. Acute and long-term effects of subthalamic nucleus stimulation in Parkinson's disease. Stereotact Funct Neurosurg 1994;62:76–84.

Brown EA, Ross JD, Blum RA, Yoonkey N, Wheeler BC, DeWeerth SP. Stimulus-artifact elimination in a multi-electrode system. IEEE Trans Biomed Circuits Syst 2008;2:10–21.

Carlson JD, Cleary DR, Cetas JS, Heinricher MM, Burchiel KD. Deep brain stimulation (DBS) does not silence neurons in subthalamic nucleus in Parkinson's patients. J Neurophysiol 2010;103:962–7.

Dostrovsky JO, Levy R, Wu JP, Hutchison WD, Tasker RR, Lozano AM. Microstimulation-induced inhibition of neuronal firing in human globus pallidus. J Neurophysiol 2000;84:570–4.

Erez Y, Czitron H, McCairn K, Belelovsky K, Bar-Gad I. Short-term depression of synaptic transmission during stimulation in the globus pallidus of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine-treated primates. J Neurosci 2009;29:7797–802.

Harding GW. A method for eliminating the stimulus artifact from digital recordings of the direct cortical response. Comput Biomed Res 1991;24:183–95.

Hashimoto T, Elder CM, Okun MS, Patrick SK, Vitek JL. Stimulation of the subthalamic nucleus changes the firing pattern of pallidal neurons. J Neurosci 2003;23:1916–23.

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.

Heffer LF, Fallon JB. A novel stimulus artifact removal technique for high-rate electrical stimulation. J Neurosci Methods 2008;170:277–84.

Merrill DR, Bikson M, Jefferys JG. Electrical stimulation of excitable tissue: design of efficacious and safe protocols. J Neurosci Methods 2005;141:171–98.
Moliadze V, Zhao Y, Eysel U, Funke K. Effect of transcranial magnetic stimulation on single-unit activity in the cat primary visual cortex. J Physiol 2003;553: 665–79.
Montgomery Jr EB, Gale JT, Huang H. Methods for isolating extracellular action potentials and removing stimulus artifacts from microelectrode recordings of neurons requiring minimal operator intervention. J Neurosci Methods 2005;144:107–25.
Moran A, Bar-Gad I. Revealing neuronal functional organization through the relation between multi-scale oscillatory extracellular signals. J Neurosci Methods 2010.
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.
Ranck Jr JB. Which elements are excited in electrical stimulation of mammalian central nervous system: a review. Brain Res 1975;98:417–40.
Strafella AP, Vanderwerf Y, Sadikot AF. Transcranial magnetic stimulation of the human motor cortex influences the neuronal activity of subthalamic nucleus. Eur J Neurosci 2004;20:2245–9.
Wagenaar DA, Madhavan R, Pine J, Potter SM. Controlling bursting in cortical cultures with closed-loop multi-electrode stimulation. J Neurosci 2005;25:680–8, %19.
Wagenaar DA, Potter SM. Real-time multi-channel stimulus artifact suppression by local curve fitting. J Neurosci Methods 2002;120:113–20.
Wichmann T. A digital averaging method for removal of stimulus artifacts in neurophysiologic experiments. J Neurosci Methods 2000;98:57–62.
Wichmann T, DeLong MR. Deep brain stimulation for neurologic and neuropsychiatric disorders. Neuron 2006;52:197–204.