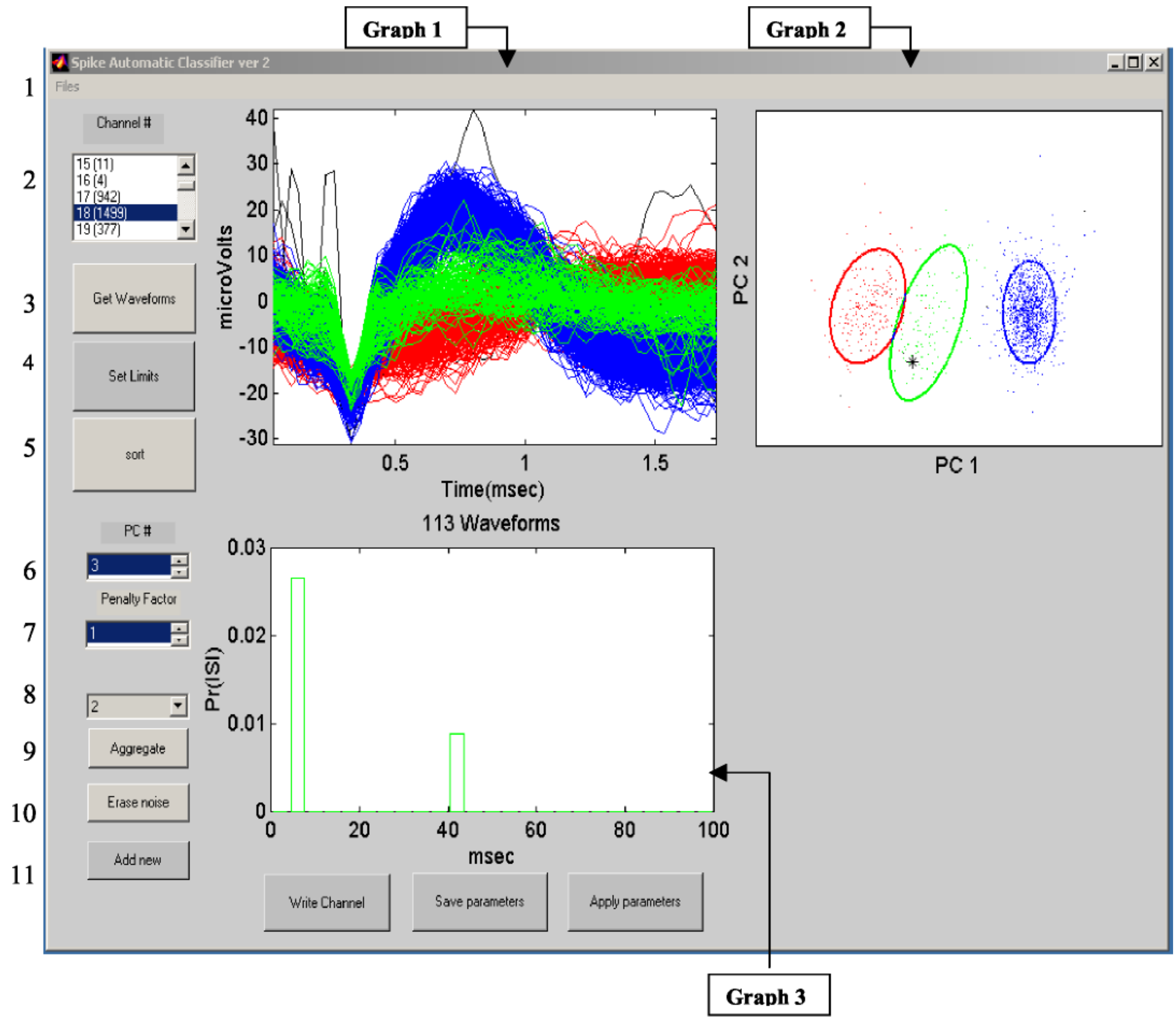
SAC (version 2) User's Guide



Introduction

SAC (Spike Automatic Classifier) is an automatic Matlab-based spike-sorting program. SAC automatically determines the number of units, and identifies the neural responses from each source out of the composite signal. It implements an agglomerative mixture decompositions algorithm based on the Expectation-Maximization (EM) algorithm, where the distribution of waveforms from each unit is modeled as a multivariate t-distribution. The algorithm automatically identifies "noise" units, but provides the user flexibility to alter the sorting parameters and sorting results. The algorithm and software were developed by Shy Shoham (Center for Neural Interfaces, Department of Bioengineering, University of Utah).

Description of buttons and functions:

1. Select one or more similarly configured *.nev or *.plx data files for analysis.
2. Select channel to analyze. This box lists all available channels and the total number of neural waveforms in each channel.
3. The "Get Waveforms" button loads a sub-sample of the total number of available waveform events, displays the waveforms in window 1 and displays the waveforms projected onto their principal components in window 2, along with an initial cluster definition. This sub-sample is used as the training data to establish the parameters for discriminating between units.
4. The "Set Limits" button produces a cross-hair cursor that is used to remove individual outlying waveforms from the sub-sample to be processed. Waveforms that fall directly under or over the cross-hair (depending on where the main group of waveforms is) are deleted.
5. The "Sort" button begins the classification process on the sub-samples. Additional options for the clustering process can be set in the PC# box (6) and the Penalty Factor box (7). The elliptical cluster definitions are continuously updated and displayed in window 2. The sorting process defines a noise cluster from random threshold crossings and overlapping spike waveforms. Noise units are displayed in black.

6. “PC#” box selection determines the number of principal components used to define each waveform in the principal component analysis. This determines the size of the feature set used in determining the clusters. However, only the first three leading principal components are used for display in graph 2. The default setting is 5.
7. A higher number selected from the “Penalty Factor” box will result in a greater number of clusters being discovered by the EM algorithm. The default setting is 1.
8. This combo box controls the inter-spike interval (ISI) histogram display, window 3. After the sorting process is completed, this selection displays the statistics of the selected unit. It also highlights the selected unit in the waveform display, window 1. Options include: noise unit, unsorted units, and the defined numbered units.
9. “Aggregrate” combines two units into one. A popup display will appear and query for the units to combine.
10. “Erase noise” removes the noise unit indicated in black.
11. “Add new” creates a new unit definition.
12. “Write Channel” classifies all the waveforms in the focus channel according to the rules determined by the training sub-sample of data. The unit definitions are saved in the currently selected data file.
13. “Save parameters” saves the parameters used for defining the units into a Matlab *.mat file.
14. “Apply parameters” loads an existing Matlab *.mat parameter file and applies the classification parameters to the current data file.

Description of Graphs:

1. Graph 1 displays the voltage-time waveforms of the neural events. Waveforms are color coded to indicate their defined units. Clicking in the middle of the display switches the view between a display of the template waveforms of the centroids, a display of a few of the waveforms and a display of all the waveforms in the sub-sample.
2. Graph 2 displays the neural waveforms projected onto their principal components. Clicking on the display switches the coordinate axes from PC2 v. PC1 to PC3 v. PC1 to PC3 v. PC2. After sorting, the memberships of the defined units are indicated by the colored ellipses, marking a distance of 2-sigma from the cluster centroid.
3. Graph 3 displays the Inter-Spike Interval (ISI) histogram of the waveforms in each unit. Button 8 controls the unit to display in this graph.

Typical Use sequence:

1. Run Sac from the Matlab command line.
2. Open one or multiple spike files (menu 1).
3. Select a channel to analyze (box 2), and load the waveforms (button 3).
4. (Optional) use button 4, and define time-amplitude windows to prune your data. 5. Sort the data.
6. (Optional) use buttons 6 & 7 to change the sorting parameters, and the re-sort.
7. Use box 8 (as well as clicking on the plots) to browse the different units. Use button 9 to aggregate units or button 11 to make the unsorted collections into a new unit.
8. When you’re satisfied with the result use button 13 to write the results into ****_ruleset.mat, and move to a different channel.

9. When you're done sorting all channels, use button 14. Choose the ruleset file and wait… All data will now be sorted.

Notes:

1. SAC can read and write to *.nev (Cyberkinetics) and *.plx (Plexon) files. A future version will probably support the NeuroShare format. Writing your own data from the Matlab workspace into a *.nev file is possible using the Matlab function write_nev which is part of the toolbox. Write_nev writes one channel at a time.
2. The initial sub-sample consists of 2000 waveforms selected from 50 evenly spaced groups in the channel. [A larger sub-sample may be warranted when you have cells that fire very sparsely – change line 115 in sac_gui accordingly].
3. Selected waveforms are screened to ensure that the waveform and the negative threshold level, set during data acquisition, actually do intersect. Waveforms that cross the positive threshold are discarded. Waveforms that saturate the A/D are discarded. If more than 70% of the channels are saturated simultaneously (amplifier "pops"), all waveform responses, in all channels, occurring at that moment are discarded.
4. Waveforms are automatically aligned so that the first peaks, after the threshold, coincide. Alignment of the peaks is accomplished by a 10x over-sampling of the peak of the waveform immediately after the threshold crossing. This portion of the waveform is fit with an interpolating spline to determine the time of the "true" peak. All waveforms are then shifted so that the spline peaks coincide and the waveform length is truncated.
5. A reduced feature set for each waveform is generated by principal component analysis and the sorting process is then applied to this reduced feature set. The Fuzzy C-Means clustering algorithm generates

the initial location of a large number of cluster centroids. Then Multivariate t-distributions are used to model the distribution of spike waveforms, employing the Mahalanobis metric in distance calculations. An EM algorithm that estimates the parameters of mixtures of multivariate t-distributions is used in the clustering process (for more details see

Shoham, PhD thesis, 2001). This algorithm will automatically determine the optimal number and location of clusters through a process involving competitive elimination of components.

Disclaimer

The SAC software is provided for full, free and open release. It is understood by the recipient/user that Shy Shoham assumes no liability for any errors contained in the code. Although this software is released without conditions or restrictions in its use, it is expected that appropriate credit be given to the author should the software be included by the recipient as an element in other product development. In case this product is used in the process of preparing a scientific manuscript, please reference Mr. Shoham's PhD dissertation ("Advances towards an implantable motor cortical interface", Dept. of Bioengineering, December 2001), or the manuscript (in preparation).

This documentation sheet was put together by Mrs. Angela Wang (Bionic Technologies LLC., Salt Lake City). The document was last revised on December 17, 2003 (S. Shoham).