MKsort Walkthrough

Introduction

This software, MKsort, allows users to import electrophysiology data and classify waveforms as coming from different units ("sort spikes"). It also provides tools for examining the stability of these isolations over time, rating the isolations, measuring spike widths, and examining the functional tuning of the sorted units.

This document provides a basic walkthrough of a full spike sorting analysis of a set of sample data, using a file mksortDemoData.nev. This guide demonstrates all the main functionality of MKsort by doing a simple spike sorting analysis from start to finish. A user may reproduce the analysis found on this page using the sample NEV file found at http://rppl.com/downloads/mksortDemoData.nev.

Extracting Data

Before performing any sorting analysis, MKsort must first extract spike data from NEV or by reapplying thresholds to NSX files and store the spike data in a series MAT files. Spike data may be extracted from multiple files and will be organized by electrode number. To begin, start the Data Import Wizard by selecting “Extract data files…” from the File menu. The window found in Figure 1 will be presented. This window allows the user to select file type, NEV or NSX and the method for choosing data for extraction after an import directory is selected.

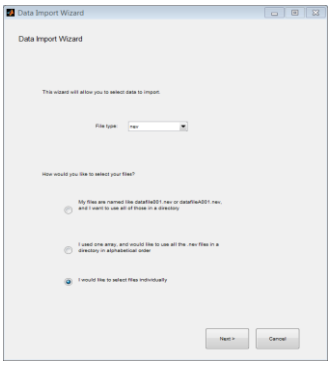

Figure 1

MKsort will allows for importing all data files from a directory or only those matching the naming scheme “datafileL###.{nev|ns5}”, where L is a single letter that identifies an array and ### represents the file number. Data files of the form datafile###.{nev|ns5} are also supported and will result in the array letter being stored as an empty string. Additionally, MKsort provides an option to select individual files from a directory.

In the next two dialogs, the user will select input and output directories. If “I would like to select files individually” is selected the user will be able to choose from NEV or NSX files found in the directory.

The final dialog is presented in Figure 2 which allows for the selection of options for data extraction. Spike waveforms may be clipped on the front or tails of the spikes to improve performance in regions of the waveforms that generally don’t have useful information. Additionally, artifact rejection may be selected and tetrode configuration files may be added. The “Conserve memory during extraction” may be needed for extremely

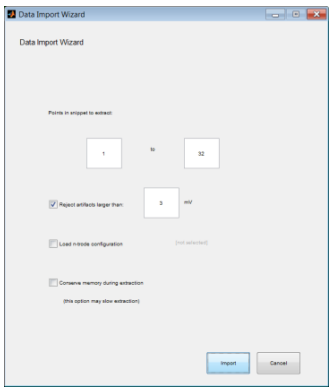

Figure 2

large files, but will significantly slow extraction and is generally not needed.

If proceeding with this walkthrough, create a directory for extracted data and use the data import wizard to extract spike data from the mksortDemoData.nev file.

Load Sorting Data

After data has been extracted, load it using the "Load sort info…" item from the File menu. MKsort data is organized by directory, so select the directory that was chosen during the data extraction process.

After loading a data directory, MKsort provides a preview of the extracted spikes by electrode. Loading the sample data should display the screen found in Figure 3. This sample data file is from a real cortical recording, but after having picked the three best spiking channels (28, 68, 95). Three poorly performing channels were also included (1, 2, 3).

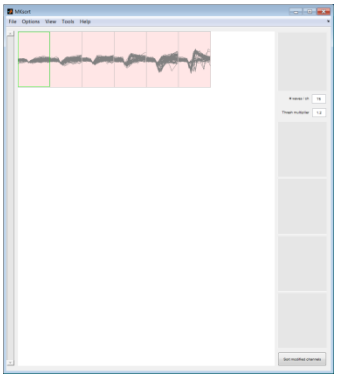

Figure 3

The preview screen allows users to browse spike waveforms and look for good channels. The user may adjust the number and the threshold of waveforms to display. The View menu provides a few options for this display, including displaying electrode IDs or sorted unit IDs.

After units are sorted, the "Sort Modified Channels" button preforms spike sorting on all channels with updated sort parameters.

Spike Sorting

To begin sorting a channel, double click on its spike preview. Figure 4, shows the Channel Sorter window for electrode 95. The "Sorting method" pull down allows the selection of either Mahalanobis templates or PCA ellipses. Both are described below.

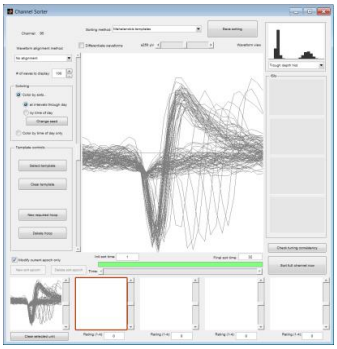

Figure 4

This spike sorter allows for sorting into up to four units. As spikes are sorted sample units are displayed in the four axes along the bottom of the Channel Sorter window and color. Unsorted units will be displayed in the lower left. Unsorted units will always be colored greyed.

Other useful options include the ability to sort based on differentiated waveforms, using the "Differentiate waveform" check, modification of the alignment of the threshold trigger, using the "Waveform alignment method" pull down, and allowing for trimming the front and back of the waveforms, using the "Init sort time" and "Final sort time" text boxes.

The slider found left of the main axis, changes the threshold to display waveforms. The slider above the main axes adjusts the scale of displayed waveforms.

Mahalanobis Templates

Template sorting is the default sorting method in the Channel Sorter window. Template sorting allows for the selection of waveforms which are then used as a sorting template. The mean waveform of those selected becomes the template. Matched waveforms are chosen by a factor of the standard deviation of each time sample (Mahaloanobis distance) of the mean waveform. In addition, hoops may be added independently of the template sorting.

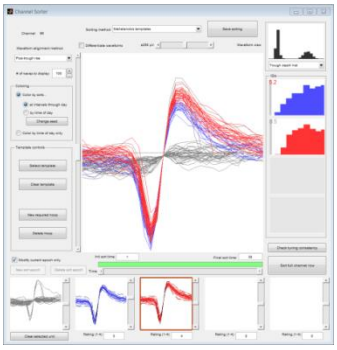

Figure 5

To begin, ensure that "Mahalanobis Templates" is selected from the "Sorting method" pull down. If following along with the walk through, ensure that Channel Sorter is opened with electrode 95.

Electrode 95 aligns well with the "Post-trough rise" alignment selected, so select this option. To begin sorting, select one of the empty squares on the bottom row of axes. As discussed above, the second axes from the right refers to unit 1.

After selecting unit 1, chose waveforms for template sorting by drawing a line on the main axes that crosses the units that are of interest. Afterward, click on the axes for unit 2 and select a different set of templates waveforms. The slider bars on the units axis increases or decreases the acceptance for sorting. If templates are selected for more than one unit, each unit will compete for waveforms. Additionally, templates may be applied to the unsorted category to help remove noise.

In Figure 5, template waveforms have been selected so that approximately half of the waveforms of the obvious unit seen on this channel have been sorted into units 1 and 2. This demonstrates the templates competing for waveforms. Adjusting the sliders on the axes for units 1 and 2 causes waveforms to be reclassified as the corresponding unit.

MKsort then allows for user defined ratings for each sorted unit between 1 and 4. 1 corresponds to a poorly defined unit and 4 being a perfectly sorted single unit. If proceeding with the walkthrough, score these two units now. The ratings will be used in later analysis performed by MKsort.

Up to now, only the previewed waveforms have been sorted so that the sorting parameter can be adjusted quickly. The "Sort full channel now" will apply the sorting parameters to all the waveforms in this channel. Figure 5 should now show the fully sorted channel and display the inter-spike interval for each channel. Click "Save sorting" to save sorting parameters and return to the main MKsort window.

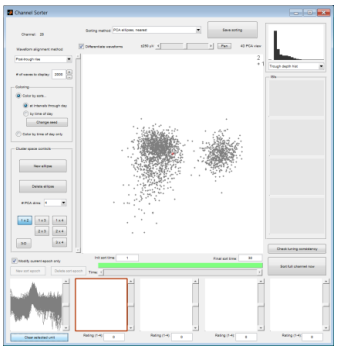

Figure 6

PCA Ellipses

To begin doing a PCA sorting analysis of an electrode, double click on the preview of the electrode's preview screen on the main MKsort window and select "PCA ellipses, nearest" from the "Sorting method" pull down menu. This walkthrough will use electrode 28 for PCA analysis.

In PCA space, electrode 28 looks good using the “Post-trough rise” for the alignment selection and “Differentiate waveforms” selection. With these selected, the user should be presented with Figure 6.

The “Cluster space controls” allows for modifications of the display and sorting using the first four PCA dimensions. The “3-D” button creates a new figure that provides a scatter plot of the first three PCA dimensions. The “Pan” button found about the PCA axis will allow the user to drag the PCA scatter plot to best show the data.

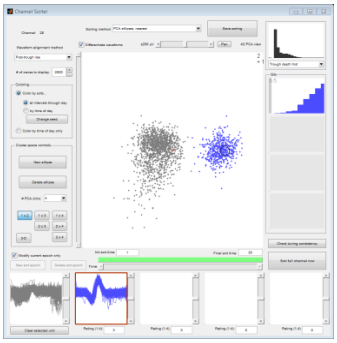

Figure 7

To begin sorting, select a “New ellipse” and position it over the right clump of waveforms. The acceptance used to sort this unit may be adjusted by reducing the size of the ellipse or adjusting the slider to the right of the unit preview axis. Give this unit a rating and click “Sort full channel.” Again a histogram of the inter-spike intervals will be displayed on the right for each sorted unit.

If following this walkthrough, a user should now see Figure 7 at this point. Click “Save sorting” to return to the main MKsort window.

Sorting by Epoch

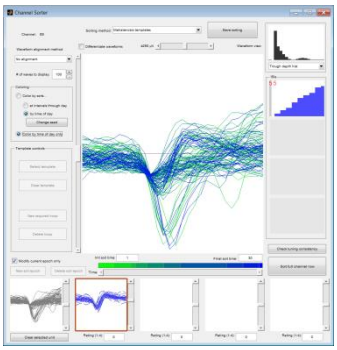

Figure 8

MKsort allows for previewed spike waveforms to be colored and sorted by time during the recording. This allows users to assess the how much a unit may be changing during the course of a recording and to adjust its sorting parameter accordingly. If following the walkthrough, open electrode 68 in the Channel Sorter.

In the “Coloring” panel, select “by time of day.” The slider below the main display will now be enabled and the previewed waveforms are displayed only if occurred at times corresponding to the slider position. Moving this slider will change the times that are display in the waveform previews.

Sorting parameters may now be adjusted by time epochs. To apply sort parameters only to particular epochs, drag the slider bar below the main display axis and click the “New sort epoch.” Sorting template selection and hoops may be adjusted independently in each epoch.

Another useful visualization tool is to color waveforms in the main display by time during the recording. In the “Coloring” panel, select “Color by time of day only”. The main display should now show colors from blue to green, with spikes early in the recording being colored blue and those later in the recording being colored green. This coloring may help the eye find drift in the previewed waveforms.

Figure 8 shows an example of an analysis on electrode 68 using different sort epochs. If continuing with the walk through, sort this channel and save the sorting parameters. Don’t forget to give the units a rating.

Trough to Peak Analysis

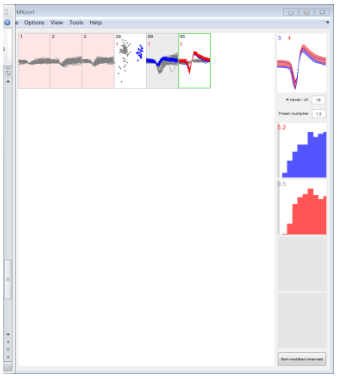

Figure 9

As sorting parameters are modified and saved for each channel, the main MKsort program keeps track of modified channels. To fully sort all channels and adjust structures in the main MKsort window, push the "Sort modified channels" button at the lower right of the MKsort window. This will update previews to include the sorted units and display inter-spike intervals and waveform envelops. If following through this walkthrough, the user should be presented with Figure 9.

After having had updated channels with modified sort parameters, Trough to Peak (TTP) analysis can be performed. From the "Tools" menu select "Calculate TTP durations." The user will then be prompted to determine the threshold for unit ratings to be used in this analysis. The default threshold is 2.1, though these ratings are arbitrary to a degree. Note that for this analysis to work, there must be units with ratings above this threshold.

Figure 10

Afterwards the TTPs are calculated and the user is presented with a window that will allow the user to manually remove units. If following the walkthrough, the user will be presented with Figure 10. In this case, we only have four units (and the two from channel 95 were the same), but in a more realistic analysis there would be more.

After choosing the interesting units, publication quality plots are produced with of the mean unit waveforms and the TTP histograms.

Adding Trial Info

MKsort includes the ability to add trial information to spike sorting. This allows the "Channel Sorter" to plot spikes rates based on sorted spikes to be plotted as a function of trial condition.

Trial info may be added from a MATLAB structure contained in a MAT file that is paired with two MATLAB functions. One function will be used to extract data from the structure. The second function will chose the events and trials where are of interest to sorting. By default, MKsort uses a structure named "R", called an "R structure". The default function for extracting data from R structures is trialInfoFromR.m. The data may be customized by creating a new structure format and corresponding function. See User Preferences Section for more details.

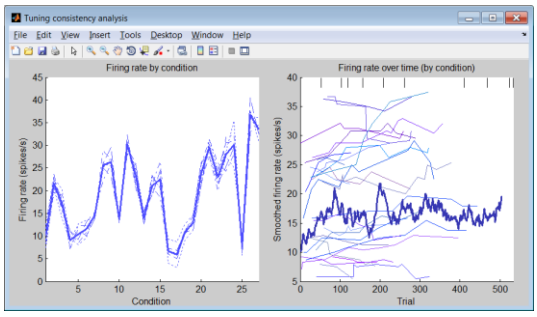

Figure 11

Select "Add Trial/Condition Info…" from the File menu. A file open dialog will be opened to select a MAT file

which contains the R structure. A dialog is presented to select the function used to select data from the extracted data. MKsort ships with a few examples of conditions functions, the simplest being condsByTrialSimple.m. Select this function by entering "condsByTrialSimple" in the text field.

To see the effect of the trial data, open Channel Sorter using electrode 68. If electrode 68 has been sorted, select "check tuning consistency" from the right side of the Channel Sorter window, just below the ISI histogram axes.

A figure similar to Figure 11 should be displayed with two axes, one containing the spike rates versus trial condition and one showing spike rates for each trial. On the left axis, in addition to plotting the spiking rate, the first few principle components are calculated and the spike waveforms are sorted by their scores in PCA space. The rates of the waveforms corresponding to the best and worst scores are overlaid on the overall firing rate to provide a measure of the uncertainty. The plot on the right shows the firing rate versus trial number. The bold line includes all conditions, and the others show the firing rate for each trial condition. In this example, there are 27 separate trial conditions, so the plot looks very busy. In real experimental data, the conditions function will determine the number of trial conditions.

The conditions function is an arbitrary function that accepts as parameters the R structure and index of the electrode array and returns three MATLAB arrays giving a list of the codes for each trial (for example digital data from an instrument) and the start and stop times in seconds for each trial. For more details, see the functions trialInfoFromR.m and condsByTrialSimple.m.

User Preferences

The file userPrefsMKsort.m is loaded by the main mksort function and there are a few options to customize the way MKsort runs. The handles.defaultDataDir will change the default directory for finding neural data and saving resulting analysis. Custom functions to extract data trial data can also be provided here by modifying the trialInfoFcn structure. The function used to copy data into the sorts structure (see Section "MAT File Structre") may also be changed here using the mergeSortsIntoDataFcn.

For more information about this file, see the MATLAB source in the files userPrefsMKSort.m and mksort.m. If creating trial data structures, see how the R structures are used in trialInfoFromR.m and mergeHandsortsIntoR.m as well as the trial condition functions.

MAT File Structure

As has been described in the Section "Extracting Data", MKsort extracts spikes from NEV or NSX files and stores it in MAT files before doing any processing. All the sorting parameters and unit identifications are then saved in MAT format.

A useful aspect of MKsort is that the sorted data and extracted spikes are easily accessible in MATLAB for customized analysis and for the production of publication quality figures. Here a brief description of the MAT file structures is provided.

Waveforms

A waveforms structure is created for each electrode found in the imported data files. They are stored in files named: waveforms_XXX.mat, where XXX corresponds to the electrode ID. A waveform structure contains the extracted data and some important information about sorted units.

Field	Description
electrode	Electrode ID
rating	User defined rating provided during sorting, ranges from 0-4. This is a subjective measure of how close the sorting is to a single unit. A perfect unit being and a very poor being 1. Zero refers to an unsorted unit.
alignedWaves	Waveforms of spikes after alignments have been applied.
alignMethodFcn	Function name of the aligned method used for this electrode if an alignment method is used.
array	Electrode array identifier as a single letter. This code is taken from the data file names. See Section "Extracting Data" for more information. NaN if filename format is not used.
sorted	Boolean. 0 – unsorted, 1 – sorted
sourceFiles	Structure containing file name and the total number of spikes in the file.
spikeTimes	Time in seconds of all spikes found in the file from the start of the file.
trialInfo	Trial info data extracted from R structure. The exact form of this will depend on the trial info functions that are used.
units	Unit numbers for sorted units in the range 0-4. Zero corresponds to unsorted units.
waves	Waveforms of all extracted spikes in units of µV.

Sorts

The sorts structure contains saved information about sorted units and sorting parameters. The sorts structure is found in the file sorts.mat.

Field	Description
differentiated	Boolean. 0 – Use spike waveform for sorting. 1 – Use the differentiated waveform for sorting.
electrode	Electrode ID
fullySorted	Boolean. 0 – Sorting parameters need to be applied to all waveforms. 1 – Sorting has been performed or no changes have been made in sorting parameters.
maxRatings	Holds the user ratings for each sorted unit for each electrode.
array	Array code as a single letter. Taken from NEV or NSX file name.
nPCADims	Number of PCA dimensions used in sorting.
onlineSorted	Boolean. Whether sorting has been applied online during recording.
autocorrs	Details of corrections made to some Cerebus files due to trigger issues.
rated	Boolean. 0 – User rating has not been applied. 1 – User rating has been applied.
sorts	Structure containing details of spike sorting parameters.
userSorted	Boolean. 0 – User rating has not been applied. 1 – User rating has been applied.

viewMode	Choice of data to display on main mksort screen. Either 'PCA' or 'Waveform'.
waveEnvolope	Top and bottom envelope for each sorted unit.

Previews

Data needed to plot previews on the main screen for fast refreshing. This structure is found in the file previews.mat.

Field	Description
electrode	Electrode ID
array	Code identifying electrode array. Taking from NEV or NSX filename.
throughDay	Subsampled waveforms for quick display on mksort screen.
throughDayAligned	Aligned version of throughDay if the channel has been sorted.
throughDayUnits	Units of throughDay waveforms, used for coloring.
thoughDaySpikeNums	Spike number of throughDay waveforms.
PCAPts	PCA data for PCA scatter plot.

Questions or Comments

Questions, comments, and bug reports are greatly appreciated. Please direct them to https://github.com/ripple-neuro/mksort/issues or send email to support@rppl.com.