Spike sorting with linear algebraic transformation of spike shapes using LabVIEW software: an application for single unit recordings

[1,2] Andriy Oliynyk, [1]Claudio Bonifazzi, [1]Benno Gesierich, [1]Livio Finos, [1]Luciano Fadiga

[1] – *Section of Human Physiology, Faculty of Medicine, D.S.B.T.A., University of Ferrara, Italy.*
[2] – *Normal Physiology Department, Odessa State Medical University, Ukraine*

Introduction and background

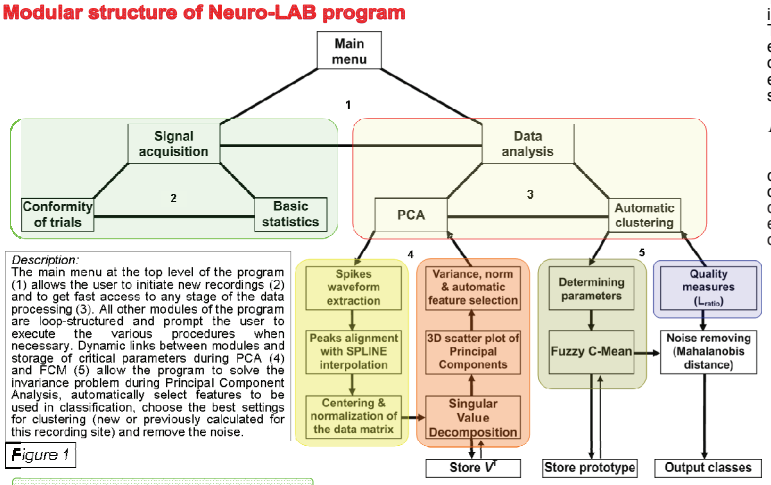
Electrophysiological recordings of single neuron activity is a fundamental tool to investigate brain functions. Although, traditionally, single neurons are isolated and discriminated by means of threshold detectors, in the majority of cases the recording electrode picks up signals from several cells from a single recording site [1]. A number of algorithms and methods have been employed to classify different neuronal discharges and to isolate single neurons from multispike recordings [4]. Advanced computational procedures may improve accuracy in multiple signals acquisition, but they are often highly interactive, requiring specific experience and arbitrary judgments. To solve this puzzling antagonism between speed of analysis and complexity of the data, we suggest a new approach to: 1) perform data acquisition; 2) extract Principal Components (PC) of spike shapes using vector-based processing of the waveform data matrix; 3) classify PC clusters by Fuzzy C-mean (FCM) algorithms applied to the multi-dimensional space.

Description of spike sorting procedure

Electrophysiological recordings

Animals: macaque monkey (*Macaca fascicularis*).
Electrodes: tungsten microelectrodes with impedance 0.15–1.5MΩ.
Amplifier: BAK Electronics (Germantown MD), USA.
Gain: x10000.
Filter: bandpass 300-6000 Hz.
Digitization: DAQ - PCI-6071E (National Instruments, USA) at 10 kHz.
Software: Neuro-LAB (UNIFE, Italy).

Modular structure of Neuro-LAB program



Description:
The main menu at the top level of the program (1) allows the user to initiate new recordings (2) and to get fast access to any stage of the data processing (3). All other modules of the program are loop-structured and prompt the user to execute the various procedures when necessary. Dynamic links between modules and storage of critical parameters during PCA (4) and FCM (5) allow the program to solve the invariance problem during Principal Component Analysis, automatically select features to be used in classification, choose the best settings for clustering (new or previously calculated for this recording site) and remove the noise.

Figure 1

Signal acquisition and statistics

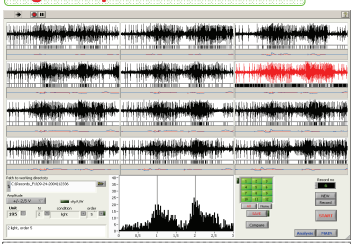


Figure 2a As soon as one experimental condition is acquired (in this case 12 trials), the basic statistical analysis on conformity of trials is performed using uniformity of neural discharges that were isolated by hardware threshold discriminator. Bad or inappropriate trials can be discarded and overwritten immediately.

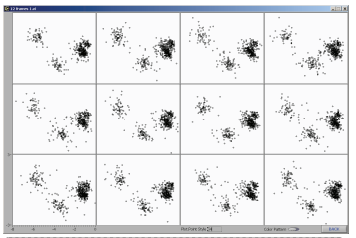


Figure 2b Neuro-LAB program permits to obtain very fast two first Principal Components (PC1 and PC2) for registered spike waveforms and to visualize them in scatterplots for each trial. The visual representation of clouds allows to evaluate acquisition stability and quality.

Spikes waveform extraction

For each action potential recorded extracellularly, the quadratic fit was tested against a threshold level, interactively adjusted for each recording site *(Fig. 3)*. Five samples before the peak and 7 samples after it (1.2 ms in total) were collected for each spike for further analysis *(Fig. 4)*. Spikes that violate a minimum refractory period by occurring within a fixed time window after the preceding threshold crossing were discarded from the analysis.

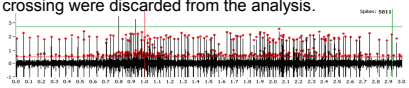


Figure 3

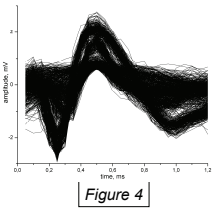
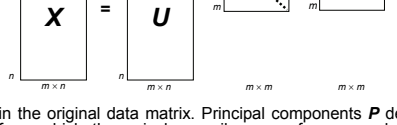


Figure 4

MATRIX DEFINITION, SINGULAR VALUE DECOMPOSITION AND PRINCIPAL COMPONENT ANALYSIS

The matrix $\boldsymbol{X}^{(nxm)}$ is a 2D data-set where indexes n specify the detected spikes and m indicates the digital samples describing each spike. This representation makes possible to apply the matrix-based procedures available in LabVIEW to center and scale the peaks of the collected spikes, as well as to perform the singular value decomposition (SVD) of matrix $\boldsymbol{X}$. SVD is a factorization procedure [6] that transforms the data matrix $\boldsymbol{X}$ into a matrix product $\boldsymbol{X}=\boldsymbol{USV}^T$, where vectors in the left, $\boldsymbol{U}^{(nxm)}$, and right, $\boldsymbol{V}^{(nxm)}$, singular matrixes are orthogonal and the diagonal matrix $\boldsymbol{S}^{(nxn)}$ collects the m singular values $s_1 \geq s_2 \geq \ldots \geq s_m \geq 0$; the superscripts $\boldsymbol{V}^T$ indicates that matrix $\boldsymbol{V}$ is transposed.



The SVD algorithm is a straightforward method to perform the Principal Component Analysis of the matrix containing spikes. Indeed, by carefully selecting the significant singular values in $\boldsymbol{S}$, the data matrix $\boldsymbol{X}$ can be decomposed into a bilinear product $\boldsymbol{TP}^T$ and a matrix $\boldsymbol{E}$ containing noise [3]:

$$X = \hat{X} + E = TP^T + E$$

where the *scores vectors* $\boldsymbol{T}=\boldsymbol{US}$, and the *Principal Components* $\boldsymbol{P}=\boldsymbol{V}$ in matrix $\hat{\mathbf{X}}$ give reason of the large amount of variance in the original data matrix. Principal components $\boldsymbol{P}$ detail the spikes profile and map vectors space from which the noiseless spikes waveforms can be extracted by a two-step procedure (*spikes sorting*). The 1st step is a spikes classification procedure based on the *clustering* of the coordinates of vectors in $\boldsymbol{T}$ along the PCs directions; the 2nd step is the spikes extraction from $\hat{\mathbf{X}}$. It is worthwhile to stress that for a given set of spike shapes (i.e., when the properties of the recording site -position of electrode, number and character activity of neurons- are stable) the scores matrix $\boldsymbol{T}$ results from the direct projections of matrix $\boldsymbol{X}$ along the principal component vectors, $\boldsymbol{T}=\boldsymbol{XP}$, without performing SVD. This projection is used during on-line procedure for monitoring the neuron activity in the PCs space.

Fuzzy C-Mean clustering

To classify the spikes in the PCs space, an iterative FCM-*clustering* algorithm based on the classical ISODATA method was used [2]. According to that procedure, the number of clusters needs to be specified beforehand. A 3D scatter plot of the first three PCs allows inspecting the spatial distribution of scores coordinates and their local density (Fig.5).

Each point in 3D PCs space represents a spike shape: 4 dense clusters are clearly visible and good separated. The point's position in PCs space was evaluated by the Mahalanobis distance and the dots far away from any cluster centers – the unclassified spike waveform – are marked as "red dots".

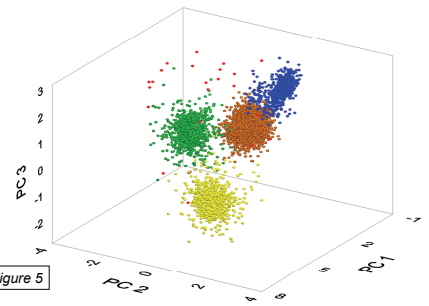


Figure 5

Measures of cluster validity

Quality of clustering is an important criterion which is based on the compactness and separation of the to-be-identified clusters. Several figures of merit, such as the partition coefficient, the partition exponent, the classification entropy, have been used to estimate the validity of clustering [2].
To quantitatively estimate the efficiency of FCM procedure we used here for clustering our data, we implemented the objective criteria known as L_{ratio} and *Isolation Distance* as part of our routine [5]. These figures of merit were used to evaluate the clustering efficiency in finding clusters of extracellularly recorded spikes in behaving animals, by calculating how well separated are the spikes of each cluster from theother spikes recorded by the electrode from the same location. We firstly evaluated the $L(C)$ value as a function of the *Isolation Distance*, $D^2_{i,C}$; i.e., the distance of isolated spikes i to cluster C:

$$L(C) = \sum_{i \notin C} 1 - CDF_{\chi^2_{df}}\left(D^2_{i,C}\right)$$; where $i \notin C$ is the set of spikes which are not members of the cluster and $CDF\chi^2_{df}$ is the cumulative distribution function of the χ^2 distribution with $df=3$. A low value of L indicates that the cluster is well separated from the surrounding spikes. The L_{ratio} is given by dividing L by the total number of points in the cluster. This ratio allows to obtain stable cluster evaluation from a particular recording site, and to take in account that clusters with a larger number of spikes better tolerate contamination from neighboring clusters.

On-line spike sorting

The spike sorting in real-time is a crucial task when the time of investigation is limited by particular situations (e.g. electrophysiological recordings in neurosurgery patients). On the basis of principles and algorithms used for off-line application we have created another subprogram (subVI) making the spike sorting on-line. The principal structure and main elements are presented in *Fig. 6*.

Description: VI 1 and VI 2 (marked by dashed line) – are two separated Virtual Instruments working simultaneously on one or two different computers connected by LAN. Solid blocks and lines inside VI 1 element are parts of the program involved in the on-line procedures; dotted blocks and lines are parts running once to get parameters before starting the on-line processing.

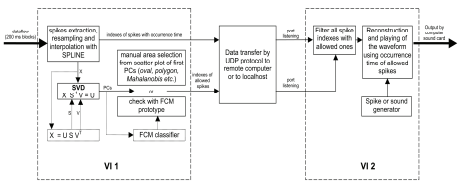


Figure 6

Example of Unit Classification

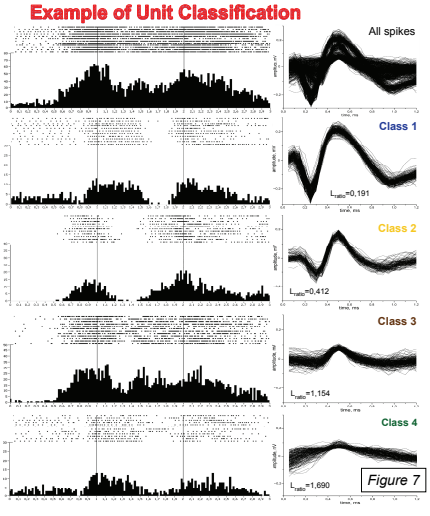


Figure 7

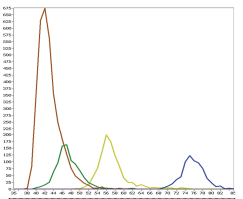


Figure 8. Norm distribution for each class separately.

The outcome of PCs successful classification can be backwards applied to the raw signals to build rasters and histograms describing individual neurons' response during the experimental task (Figure 7). Given example shows two neurons close to the electrode showing a reciprocal electrophysiological behavior (Class 1 and Class 2). The neuron isolated as Class 3, despite its low amplitude of discharge, is still well separated by our technique, notwithstanding the fact that it is greatly overlapping with noise. The separation from noise becomes dramatically efficient in the case of Class 4 neuron (see Figure 8).

Conclusion

1. Our procedure allows to automatically classify a high number of individual neurons from each recording site without requiring a significant interaction with the experimenters. Quantitative parameters of clusters quality showed excellent efficacy to discriminate different neurons even when they show a discharge pattern only marginally higher than recording noise.
2. The intelligible and user-friendly graphical structure of our software renders it handy also for the clinical uses (i.e. during neurosurgery recordings).
3. Our way of implementation of spike sorting algorithm can be easily integrated in the LabVIEW environment. The LabVIEW flexible ways of data visualization, including interactive 3D representation of results, allows fast visualizing the neuronal clusters and easily correlating neural responses to behaviorally relevant epochs.

References

1. Abeles M. A journey into the brain. In: *Signal Analysis and Pattern Recognition in Biomedical Engineering*, edited by Inbar G. New-York: John Wiley & Sons, 1974, p. 41-59.
2. Bezdek J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.
3. Jackson, JE. A user's guide to principal components. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, New York, 1991.
4. Lewicki MS. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network* 9: R53-78, 1998.
5. Schmitzer-Torbert, N., Jackson, J., Henze, D., Harris, K., Redish, A. D. Quantitative measures of cluster quality for use in extracellular recordings. *Neuroscience*. 2005; 131(1):p.1-11.
6. Wall ME, Rechtsteiner A and Rocha LM. Singular Value Decomposition and Principal Component Analysis. In: *Practical Approach to Microarray Data Analysis*, edited by Berrar DP, Dubitzky W and Granzow M. Boston/Dordrecht/London: Kluwer Academic Publishers, 2002, p. 91-109.