# SPIKE DETECTION ALGORITHM IMPROVEMENT, SPIKE WAVEFORMS PROJECTIONS WITH PCA AND HERARCHICAL CLASSIFICATION

**E. Biffi *, D. Ghezzi [#], A. Pedrocchi [†], G. Ferrigno [†]**

* Politecnico di Milano, Bioengineering Department, Neuroengineering and Medical Robotics Laboratory, p.zza Leonardo da Vinci 32, 20133 Milano, Italy. emilia.biffi@mail.polimi.it

[#] Neuroscience and Brain Technologies Department, The Italian Institute of Technology, via Morego 30, 16163 Genova, Italy

[†] Politecnico di Milano, Bioengineering Department, Neuroengineering and Medical Robotics Laboratory, p.zza Leonardo da Vinci 32, 20133 Milano, Italy.

**Keywords:** neuroengineering, MEA, PCA, cluster analysis.

## Abstract

Definition of single spikes from multiunit spike trains plays a critical role in neurophysiology and in neuroengineering. Moreover, long period analysis are needed to study synaptic plasticity effects and observe the long and medium term development on which all Central Nervous System (CNS) learning functions are based. Therefore, the increasing importance of long period recordings makes necessary on-line and real time analysis, memory use optimization and data transmission rate improvement. A threshold-amplitude spikes detection method is chosen and 5 noise level estimate methods were developed. Than APs are bundled to group using principal component analysis and classified (hierarchical classifier). The system has lot of applications like high-throughput pharmacological screening and monitoring effects.

## 1 Introduction

Neuron is the most important elaboration unit in the Central Nervous System (CNS).

In the last decade, studies on neuronal physiology and plasticity have provided a detailed picture of the molecular mechanisms underlying the modulation of neuronal activity. On the contrary, the functional mechanisms which control the network properties remain poorly understood, and represent a new frontier in the neuroscience [10].

The majority of the neurons of the CNS communicate by means of electric pulses, called action potentials. These brief voltage spikes can be recorded with conventional electrodes, such as micropipettes, which are able to acquire also the signals of a single neuron. On the other side, the number of simultaneous measurements is limited by the employment of micropipettes, because of the difficult placement of more than two micropipettes on the same slide [7].

New technologies have been developed in order to overcome the experimental difficulties which limit the number of neurons studied at the same time, the temporal constraints due to the placement of the electrodes and the mechanical damage of the cells during the functional study. In this context, MicroElectrode Arrays (MEA) have been designed [6]. Planar MEA, like electrodes matrixes, have been spread for the first time in the 1970s. By the means of MEA, it can be possible to study the properties of a complex neural network, not only the signals of a single neuron. These devices, produced by means of photolithographic techniques, are slides which integrate 60 microelectrodes on the surface.

Electrodes and neurons are not coupled, also in low density neuronal culture, because of MEA spatial resolution (100µm – 200µm), but each electrode records the electric potential variations in the local region, where generally there are a lot of neurons. Therefore, the detection of neural spike activity using a MEA is a technical challenge in order to study the main brain functions and to discover the functional structure of the neuronal network, i.e. how the neurons are connected and organized. Indeed, it is a primary step in order to obtain an appropriate neural network model, which represents the robustness, the versatility, and the sensitive responses of the mammalian nervous system [4, 5].

To approach this goal, it is necessary to distinguish the activity of each neuron; this means to detect the neural signal from background noise and, then, to recognise the different sources.

Moreover, long period analysis are needed to study synaptic plasticity effects and observe the long and medium term development on which all CNS learning functions are based. Therefore, the increasing importance of long period recordings makes necessary on-line and real time analysis, memory use optimization and data transmission rate improvement.

The present study can be included into this research field; it deals with the design of software for the on-line detection and classification of neural spikes, able to distinguish the activity of single cells from the signals recorded by MEA.

The system has lot of applications, from high-throughput pharmacological screening, towards neuropharmacology and monitoring effects of drugs and toxins (neurotoxicity), to staminal cells excitability researches.

## 2 Methods

An amplitude-threshold spikes detection algorithm and a hierarchical classifier were developed in Matlab 7.1 (The Mathworks, Natick, MA) environment.

### 2.1 Spikes detection

In literature there are many algorithms proposed to detect spikes from signals recorded by MEA electrodes.

The spikes detection algorithm here described is structured as follows: first the signal is band-pass filtered with a $2^{nd}$ order Butterworth filter (150Hz - 2500Hz); then, both positive and negative threshold values are determined as a multiple of noise level estimate. Finally, positive APs over positive threshold and negative APs under negative threshold are detected and validated to prevent double detections of unitary events. A detected peak is tested in ±1ms window, checking if it is the highest peak of either polarity and if the 50% of its amplitude is higher than other peak of the same polarity.

A significant improvement suggested in this work, is a qualitative evaluation of APs amplitude; spike detection algorithm make out 3 threshold values (i.e. two bits output), multiplying the noise level estimated with 3 factors. The smallest threshold not only discriminates between APs and background noise, but also identifies little peaks. Otherwise medium and high thresholds (multiples factors equal to 7 and 10) recognize middle and tall peaks.

Five noise level estimation algorithms, some suggested by literature [8, 9], some modified were developed. Then, their performances were compared with statistical analysis, to select the best solution.

The noise level can be calculated as follows.

(a) BandFlt, based on the algorithm described in [8, 9], detects spikes if the filtered stream exceeds a given multiple (supposed equal to 4) of the estimated Root Mean Square (RMS) noise. Three hundred 10ms windows of data are read; for each of these windows the RMS values are calculated. The results are sorted, and the final estimate of noise level is the 25th percentile of the values collected. Positive and negative threshold values are equal and fixed.

(b) Limada, described in [8, 9] like BandFlt, splits the data stream into 10ms windows, and determines the $2^{nd}$ ($V_{.02}$) and $30^{th}$ ($V_{.30}$) percentiles of the distribution of voltages found in each window. Then, it performs two tests: it checks if the ratio between $V_{.02}$ and $V_{.30}$ is smaller than 5 (i.e. no actual spike in the window) and if the absolute value of $V_{.30}$ is significantly non-zero (i.e. data in the window are not blanked out). If both tests are passed, the window is considered 'clean'. Noise level initialization value is defined after that 100 clean windows are collected; then the current noise level estimate is updated as described in Equation (1):

$$noise\_est(k) = 0.99*noise\_est(k-1) + 0.01*|V_{.02}| \qquad (1)$$

where k is the current iteration index.

Positive and negative threshold values are computed by multiplying noise level estimate by 4. Their values are equal and adaptive.

(c) Another adaptive algorithm suggested by literature [8, 9] and developed in this work is AdaFlt; it divides signal into 10ms windows and 128 windows of data are read. For each of these windows, the maximum and minimum values are measured; then, results are sorted, and 40th percentile of both collections is computed ($M_{.4}$ and $m_{.4}$). The noise level initialization estimates for upward ($noise\_est_p$) and downward spikes are based on the result. Thresholds are computed by multiplying noise level estimate by 2. While running, AdaFlt keeps collecting minima and maxima in 10ms windows, although it uses only one in ten windows. Whenever 128 windows have been collected, positive and negative thresholds are updated as shown in Equation (2) and Equation (3), respectively

$$noise\_est_p(k) = 0.9*noise\_est_p(k-1) + 0.1*M_{.4} \qquad (2)$$

$$noise\_est_n(k) = 0.9*noise\_est_n(k-1) + 0.1*m_{.4}\,. \qquad (3)$$

(d) AdaFlt was modified into AdaFlt 128: the former adapts thresholds after 1280 windows, the latter after 128. Therefore AdaFlt 128 is faster and more adaptive than AdaFlt, but it suffers more for noise fluctuations.

(e) The last algorithm developed in this work is Band Flt VHDL, devised as BandFlt adaptive properties improvement. It splits the data stream into 10ms windows, calculates the RMS values for each of the initial 100 windows and determines the $25^{th}$ percentile of RMS distribution ($N_{.25}$). Noise level initialization value is defined. Then the current noise level estimate is updated as described in Equation (4):

$$noise\_est(k) = 0.9*noise\_est(k-1) + 0.1*N_{.25}\,. \qquad (4)$$

Algorithms performances were statistically evaluated using a simulated neuronal signal. The simulated signal was artificially made up of both positive and negative triangular waves (each wave is composed by 20 samples); the intensity peaks range from 40 to 100µm. Apart from these simple waves, the simulated signal is characterised by complex waves, formed by 2 overlapped and temporally shifted triangular waves, representing multi-shaped AP in neuronal culture activity. The full amount of spikes that algorithm should detect of simple and complex waves, representing valid spikes, is 600.

A white Gaussian noise was used to simulate real background noise. First, it was band-pass filtered (2nd order Butterworth) between 150Hz and 2.5kHz, then normalized and overlapped to simulated signal. The Signal to Noise Ratio (SNR) is equal to 5dB, like real neuronal culture signals SNR.

Algorithms performances were evaluated performing a Screening Test that determines true and false positive (TP/FP) and true and false negative (TN/FN) values. Sensibility (Se), specificity (Sp), positive predicted value (PPV) and negative predicted value (NPV) were computed by means of Screening Test results.

### 2.2 Spikes classification

After detecting temporal occurrences of APs, validated waveforms must be extracted from signal in 2ms windows and then classified. The background assumption is that neurons usually generate APs with a characteristic shape; so the aim of classification is correlating source with a characteristic waveform.
The classification algorithm proposed in this work is based on principal component analysis (PCA) and hierarchical classification. The former finds an ordered set of orthogonal basis vectors that capture the directions in the data of largest variation [2]. The latter is a method of cluster analysis; it splits N observations (i.e. neuronal waveforms) into a series of m clusters, where m can range from 1 (all observations grouped into one cluster) to N (each observation is a cluster). The strength of this technique is that it provides the possibility of increasing or decreasing the number of clusters depending on the required level of aggregation [3].

## 3 Results

### 3.1 Spikes detection

Algorithms Screening Tests were performed; results are reported in Table 1.
It's known by literature that BandFlt underestimates noise level; therefore it identifies a lot of FP. On the other hand it produces noise level values more stable than standard deviation [8, 9].
Limada is more reliable than BandFlt in order to identify the threshold. It is an adaptive method, necessary for real time analysis. The main disadvantage is that it needs approximately 5s of signal to initialize threshold and computes a lot of FN.
AdaFlt finds a lot of FN caused by threshold excessively slow variation. On the other hand AdaFlt 128 is faster and more adaptive than AdaFlt; it identifies a few number of FN caused by noise fluctuations. Finally, BandFlt VHDL method is like to BandFlt but it is adaptive and finds less FP than BandFlt.

| | Band Flt | Limada | AdaFlt | AdaFlt 128 | Band FltVhdl |
|---|---|---|---|---|---|
| TP | 598 | 583 | 524 | 584 | 599 |
| FP | 85 | 1 | 2 | 1 | 67 |
| TN | 599315 | 599399 | 599398 | 599399 | 599333 |
| FN | 2 | 17 | 76 | 16 | 1 |

Table 1: This table reports algorithms Screening Test results.

Sensibility (Se), specificity (Sp), positive predicted value (PPV) and negative predicted value (NPV) were computed by means of Screening Test results. In this work only Se (Figure 1) and PPV (Figure 2) evaluations are described, because of other parameters not significant statistics.

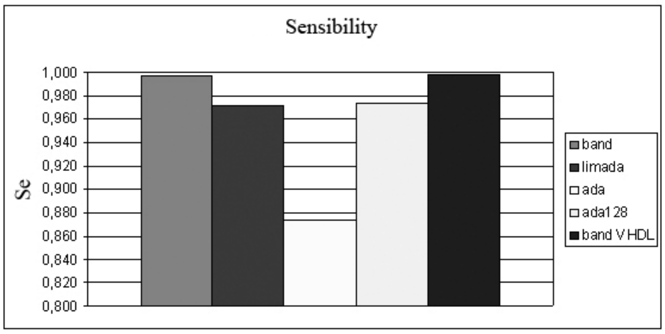


Figure 1: Here algorithms sensibility values are compared.

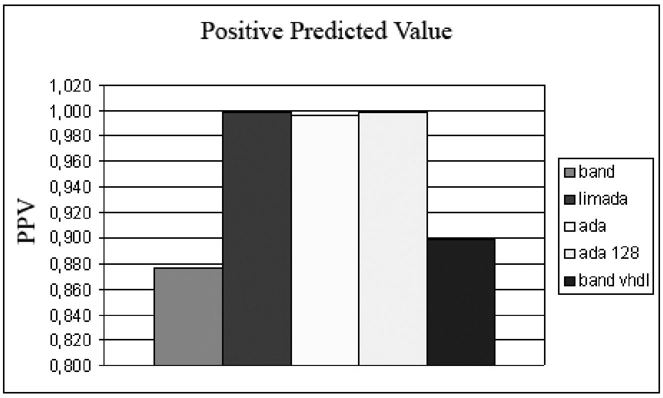


Figure 2: This figure shows the evaluation of positive predicted values between algorithms.

### 3.2 Spikes classification

To get APs separated groups, the data are projected along the first two principal components (Figure 3); it's known from literature [1, 2] that 1st and 2nd component eigenvalues can represent, alone, the useful characteristics for classification. Then, hierarchical classifier yields an accurate classification (Figure 4). The classification is satisfying: every cluster contains at least two points and distances between each barycentre are bigger than statistical data dispersion.

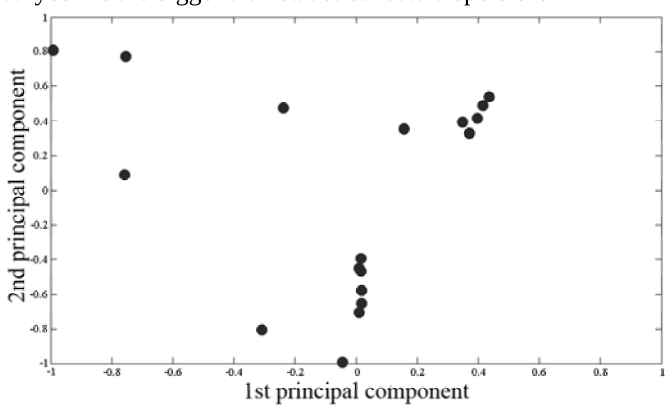


Figure 3: Data projected along the first two principal components.

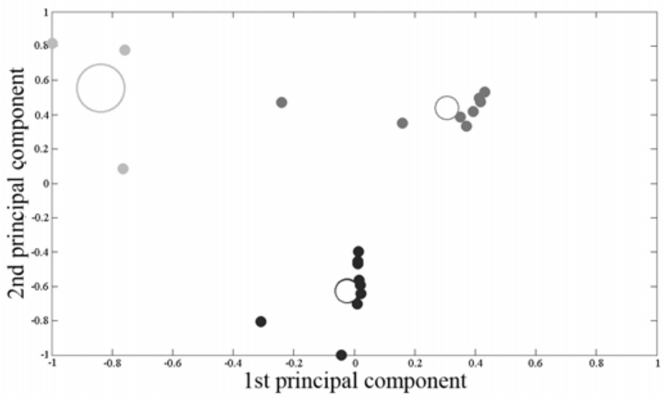

Figure 4: Hierarchical classifier output.

## 4 Conclusions

Definition of single spikes from multiunit spike trains plays a critical role in neurophysiology and in neuroengineering; indeed, it permits to understand how much information is encoded by single neurons in a neuronal network. Moreover, the possibility to develop a bidirectional communication between electronic devices and neuronal networks provides great perspectives in neuroengineering. Traditionally, the functional properties of neurons and neuronal networks have been investigated using conventional electrodes, such as glass micropipettes, thus allowing neurophysiologists to disclose a detailed picture about the single cell properties. Thirty years ago Micro-Electrode Array devices (MEAs) have been developed as tools providing distributed information about learning, memory and information processing in a cultured neuronal network. Recent applications of these technologies, above all long period analysis, have the problem of the recording and storage of the huge amount of data processed.

A threshold-amplitude spikes detection method is chosen to make software simpler allowing on-line detection and real time analysis. 5 noise level estimate methods were developed.

Band Flt finds the most number of FP but not too much FN; it has high sensibility but the PPV means a low probability of an AP correct detection. Therefore, this method is not very selective but true spikes aren’t lost. The same consideration could be written about Band Flt VHDL.

On the other hand, Limada finds the most number of FN, because it uses part of the signal to compute the initialised thresholds, FP value is very small, sensibility and PPV are very high. It means that algorithm is very selective but huge amount of data could be lost. The same consideration could be written about AdaFlt 128.

BandFlt VHDL has the highest performances in order to do long period recordings of neuronal cultures, allowing on-line and real time analysis and optimizing memory occupation.

After spike detection, the software described above, bundles to group (PCA) and classifies APs (hierarchical classifier). The classification is satisfying.

The system has lot of applications, from high-throughput pharmacological screening, towards neuropharmacology and monitoring effects of drugs and toxins (neurotoxicity), to staminal cells excitability researches.

## Acknowledgment

This work was carried out in collaboration with the ALEMBIC facility (Advanced Light and Electron Microscopy Bio Imaging Center) of the San Raffaele Scientific Institute.

## References

[1] A. Daffertshofer, J.C. Lamoth, O.G. Meijer and P.J. Beek. “PCA in studying coordination and variability: a tutorial”, *Clinical Biomechanics*, **19**, pp. 415-428, (2004).

[2] M.S. Lewicki. “A review of methods for spike sorting: the detection and classification of neural action potentials”, *Network*, **9**, pp. R53-R78, (1998).

[3] L. MacNabb. "Application of cluster analysis towards the development of health region peer groups”, *Proceedings of the Survey Methods Section*, pp. 85-90, (2003).

[4] F.O. Morin, Y. Takamura and E. Tamiya. “Investigating neuronal activity with planar microelectrode arrays: achievements and new perspectives”, *Journal of bioscience and bioengineering*, **100**, pp. 131-143, (2005).

[5] M.A. Nicolelis and S. Ribeiro. “Multielectrode recordings: the next step”, *Current Opinion in Neurobiology*, **12**, pp. 602-606, (2002).

[6] Y. Perelman and R. Ginosar. “An integrated system for multichannel neuronal recording with spike/LFP separation, integrated A/D conversion and threshold detection”, *IEEE transactions on bio-medical engineering*, **54**, pp. 130-137, (2007).

[7] P.H. Thakur, L. Hanzhang, S.H. Steven and O.J. Kenneth. “Automated optimal detection and classification of neural action potentials in extra-cellular recordings”, *Journal of Neuroscience Methods*, **162**, pp. 364 -376, (2007).

[8] D.A. Wagenaar "A versatile all-channel stimulator for electrode arrays, with real-time control", *Journal of neural engineering*, **1**, pp. 39-45, (2004).

[9] D.A. Wagenaar, T.B. DeMarse and S.M. Potter. “MEABench: A toolset for multi-electrode data acquisition and on-line analysis”, *Proceedings of 2nd International IEEE EMBS Conference on Neural Engineering*, **1**, pp. 518-521, (2005).

[10] X. Yang. “Detection and classification of neural signals and identification of neural networks” [dissertation], University of Maryland, (1989).