ELSEVIER

Journal of Neuroscience Methods 85 (1998) 81–87

JOURNAL OF NEUROSCIENCE METHODS

Rough annealing by two-step clustering, with application to neuronal signals

Pasquale Gurzi [a], Francesco Masulli [b], Arnaldo Spalvieri [c], Maria Luisa Sotgiu [a], Gabriele Biella [a,d,*]

[a] *Istituto di Neuroscienze e Bioimmagini, CNR—Via Fratelli Cervi, 93 I-20090 Segrate, Milano, Italy*
[b] *Istituto Nazionale per la Fisica della Materia and Dipartimento di Informatica e Scienze dell'Informazione, Università di Genova, Genova, Italy*
[c] *Dipartimento di Elettronica e Informazione, Politecnico di Milano, Milano, Italy*
[d] *Cattedra VI di Neurologia, Istituto Scientifico H San Raffaele, Università di Milano, Milano, Italy*

Received 31 December 1997; received in revised form 9 July 1998; accepted 12 July 1998

Abstract

To accomplish analyses on the properties of neuronal populations it is mandatory that each unit activity is identified within the overall noise background and the other unit signals merged in the same trace. The problem, addressed as a clustering one, is particularly difficult as no assumption can be made on the prior data distribution. We propose an algorithm that achieves this goal by a two-phase agglomerative hierarchical clustering. First, an inflated estimation (overly) of the number of clusters is cast down and, by a maximum entropy principle (MEP) approach, is made to collapse towards an arrangement near natural ones. In the second step consecutive partitions are created by merging, two at time previously aggregated partitions, according to similarity criteria, in order to reveal a cluster solution. The procedure makes no assumptions about data distributions and guarantees high robustness with respect to noise. An application on real data out of multiple unit recordings from spinal cord neurons of mixed gas-anaesthetized rats is presented. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Clustering; Annealing; Gibbs distribution; Hierarchical fitting; Multiple unit recordings; Rats

1. Introduction

Extracellular recordings do not often allow the clear selection of one single, well-distinguishable, spike form. The usual recordings may include more than one concurrent neuronal signal, merged into a single trace on the background noise. Thus it is necessary to implement on- or off-line techniques to identify homogeneous structures (spikes from a single neuron) in the data set (recording trace from an electrode).

Several methods proposed in the past discriminate homogeneous spike shapes without the aid of a labeled training set. The absence of category labels, i.e. markers assigned to objects, distinguishes cluster analysis from discriminant analysis. Parametric and nonparametric methods are the two fundamental approaches to clustering. In the parametric approach the distribution of data is expressed by a statistical model (e.g. a mixture of normal distributions). In the non-parametric approach the valleys of the density functions are considered as the natural boundary separating the modes of the distributions. A typical technique to construct such density functions is based on the Parzen method (Parzen, 1962).

The outcome of a crude application of the unsupervised algorithms to the separation of spikes from a neuronal recording trace is sometimes unsatisfactory. Kleinfeld and co-workers (Mitra et al., 1995; Fee et al., 1996b) recently showed that spike waveforms are highly anisotropic objects, and that their deviations are not grasped by Gaussian approaches; this has led to the

* Corresponding author. Tel.: +39 2 21717530; fax: +39 2 21717558; e-mail: biella@inb.mi.cnr.it

0165-0270/98/$ - see front matter © 1998 Elsevier Science B.V. All rights reserved.
PII S0165-0270(98)00120-4

discarding of parametric methods. A detailed guideline to this problem has been proposed by Fee et al. (1996a), a further extension to a probabilistic framework has been made by Andersen's group (Sahani et al., 1997).

The maximum entropy principle (MEP), proposed in the 50-ties by Jaynes (1957a,b), maximizes the information content of a data set. Recently some important results about the application of MEP to the clustering problem were published (Rose et al., 1990, 1993; Hofmann and Buhumann, 1997; Masulli et al., 1997). These papers describe the method of deterministic annealing applied to clustering. This approach has the nice feature of revealing the 'natural' clusters that exploit the well-known robustness of maximum entropy inference, through minimization heuristics named deterministic annealing.

In order to obtain the correct 'natural' clusters, the proper scheduling procedure for deterministic annealing must be selected. The selection of such a scheduling is often a time-demanding task. Moreover, due to the variability of the recorded signals, the scheduling must be continually reestablished at each recording session. In this paper we present a more parsimonious solution, using heuristics on the recording procedure; the solution is based on the assumption that no more than a fixed number Γ of clusters is present in our trace.

This assumption suggests to consider an overly of the true number of clusters (three times Γ, for example) and uses a tentative scheduling on a MEP-based algorithm to find a first reduced partition. Then, by an agglomerative hierarchical clustering scheme we join single partitions, two at time, until Γ is reached.

The criterion adopted in evaluating the connection strength between couples of partitions is the one used in the algorithm CLUSTER (Dubes and Jain, 1976). The algorithm of Rose et al. (1990) has been used for the MEP based algorithm.

This hybrid structure of algorithms has proved to be considerably faster and less sensitive to noise than the pure algorithm. All the algorithms were realized in MatLab language, building an easy interactive tool.

2. Data collection

The procedure of animal operation and the recording technique have been described in detail elsewhere (Biella et al., 1997).

Briefly, extracellular recordings were performed on Sprague–Dawley rats. For the experiments the rats were pre-anaesthetized with pentobarbital (40 mg/kg, i.p. for induction). A gaseous mixture anaesthesia isofluorane oxygen was then used during the experiment. The animals were paralysed with pancuronium bromide (0.5 mg/kg per h i.v.) and the dura mater opened, to allow single unit recordings from the dorsal horn of the spinal cord. Throughout the experiments, body temperature was maintained at 38°C and end-tidal CO_2 was maintained at 3.5–4%. The adequacy of anaesthesia was monitored by observing the stability of arterial pressure (values of 90–110 mmHg) and heart rate (values of 320 ± 15 beats/min).

Recording bundles of three to five tungsten electrodes were used (each electrode with a tip impedance of 0.8–1.2 MΩ). Signals were band pass filtered between 300 and 10^4 Hz. The bundles (inter-tip distances of 100 μm, in a comb array) were positioned under a dissecting microscope on the surface of the lumbar segment of the cord and advanced with 2-μm steps by an electronic micromanipulator. The electrodes were placed with the coplanar tips touching the surface of the cord. The recording sites were located at 450–800 μm corresponding to the V lamina. The depth of the cells was taken from the mean microdrive readings, noted on descent and then on ascent toward the cord surface. There was a mean difference of only 6–10 μm. In some experiments a current was passed to get a histological control of the electrode location. In all cases the microdrive readings were confirmed.

Once a number of neurons was isolated, a recording session was started. In some trials sensory stimuli were delivered to induce dynamic modifications in the spontaneous activity of the neurons. This procedure was chosen so as to clarify whether or not neuron discharges could be followed, even in the activated state. We have set the thresholds in a way that must of the data gathered is generated by multiunit activity.

All the neurons were recorded extracellularly, and their activity was acquired by the application of Schmitt-triggers with high and low pass levels suitably adapted to the noise consistency and the signal-to-noise ratio. All the spikes sampled in the simultaneously recorded trace were digitized at 25 kHz and stored together their time stamp. Sixteen time points have been collected for each spike detected; these points have been concentrated in the rise of the spike to better represent the waveforms where most of the energy is found. Data have been synchronized with respect to positive threshold crossing. Before any analysis, each spike undergoes a normalization to the [0,1] interval. On-line analysis was carried out by the appropriate programs (Data Wave, CO; Computerscope, RC Electronics, CA).

3. Maximum entropy-based clustering

In what follows x denotes the feature vector describing the spike recorded, $x \in \Re^m$, while j is the cluster index taking its values from $\mathcal{J} = \{1,2,\ldots,C\}$. Let $\Omega = \{x_k \in [1,N]\}$ be a set containing N unlabeled samples

and $\mathcal{Y} = \{y_1, y_2, \ldots, y_C\}$, $y_j \in \Re^m$, $C > \Gamma$ the set of prototypes (the spike models we are looking for).

Many classical approaches to clustering, e.g. k-means, fuzzy k-means (Duda and Hart, 1973; Bezdek, 1981), are based on the (constrained) minimization of a functional. For example, the fuzzy k-means algorithm is based on the minimization of

$$\langle E \rangle = \frac{1}{N} \sum_{j=1}^{C} \sum_{k=1}^{N} P(j \mid x_k) E(y_i, x_k) \tag{1}$$

under the constraint

$$\sum_{j=1}^{C} P(j \mid x_k) = 1, \forall x_k \in \Omega. \tag{2}$$

In Eqs. (1) and (2) $E(y_j, x_k)$ is a dissimilarity measure (distance or cost) between a sample x_k and the prototype y_j of a specific cluster j, and $P(j \mid x_k)$ the conditional probability of the specific cluster j. Often $E(y_j, x_k)$ is assumed to be the square of the Euclidean distance, i.e.

$$E(y_j, x_k) = \|x_k - y_j\|^2. \tag{3}$$

Pitfalls linked to this approach are: (1) the possibility of getting stuck in local minima of $\langle E \rangle$; and (2) the necessity to assume a value for C.

Some researchers have proposed a clustering method based on the maximum entropy principle and a deterministic annealing procedure (Rose et al., 1990, 1993).

This method assumes that the conditional probabilities are Gibbs distributions:

$$P(j \mid x_k) = \frac{\exp(-\beta E(y_j, x_k))}{Z_k}, \tag{4}$$

where

$$Z_k = \sum_{j=1}^{C} \exp(-\beta E(y_j, x_k)) \tag{5}$$

is a normalization factor named partition function (Rose et al., 1990). From a statistical mechanics point of view, the Lagrange multiplier β is interpreted as the inverse of temperature T ($\beta = 1/T$).

Moreover, β can be interpreted as a control parameter of fuzziness. When β increases, the association of samples with clusters becomes sharper. As a consequence, the probability $P(j \mid x_k)$ can be interpreted as the membership function of the pattern to the cluster, as in the fuzzy clustering theory (Bezdek, 1981).

The limit cases are:

- for $\beta \to 0^+$, we have $P(j \mid x_k) = 1/C$ for all j, k, i.e. each sample is equally associated with each cluster;
- for $\beta \to +\infty$, we have

$$P(j \mid x_k) = \begin{cases} 1 & \text{for } j = \operatorname{argmin} E(y_j, x_k), \forall_j \mathcal{J} \\ 0 & \text{elsewhere} \end{cases},$$

i.e. each sample is associated with only one cluster (hard limit), as in the k-means approach.

Let us define the effective error (also named the free energy, by analogy to statistical mechanics)

$$F = \frac{1}{\beta} \ln Z, \tag{6}$$

where $Z = \Pi_k Z_k$ is named the total partition function.

A relevant property of F for the clustering problem is that

$$\lim_{\beta \to \infty} F = \langle E \rangle. \tag{7}$$

This limit allows one to find the solution of the constrained minimization of $\langle E \rangle$ (i.e. the solution of the k-means) by performing a deterministic annealing on F, as has been proposed (Rose et al., 1990, 1993). In the algorithm used here $E(y_j, x_k) = \|x_k - y_j\|^2$ is assumed; then, following a temperature scheduling, a minimization of F with respect to $\mathcal{Y}$ is performed for each value of T by iterating the following formula (Rose et al., 1990):

$$\hat{y}_j = \frac{\sum_{k=1}^{N} P(j \mid x_k) x_k}{\sum_{k=1}^{N} P(j \mid x_k)} \text{ for all } j \tag{8}$$

with $P(j \mid x_k)$ given by Eq. (4). The resulting MEP ALGORITHM is:

```
c ← overly (tree time Γ, for example)
initialize Y randomly on Ω;
test ← 1;
S ← δ; (arbitrary small positive value)
β ← β_min;
J ← [1, 2,…,C];
WHILE β < β_max
   WHILE test > S


   ŷ ← Eq. (8);
   test ← max_{j∈J} (E(y_j, ŷ_j));   y ← ŷ;
END
test ← 1;
β ← β + step;
END
```

It is worth noting that, whereas many well known (k-means, fuzzy k-means) clustering algorithms need the specification of the number of clusters, this al-

gorithm can start with an over-dimensioned number of clusters (we typically use $C = 3\Gamma$).

In fact, at high temperatures, all the prototypes collapse at the one point, and then, during annealing, 'natural' clusters differentiate. As mentioned in the introduction, the selection of the scheduling strongly influences the results. A consequence of the use of the wrong annealing scheme is a differentiation excess of the clusters. In order to speed up the analysis procedure, we used a rough scheduling procedure to find a first partition, then further optimized this partition with an agglomerative hierarchical clustering scheme that will be shown in the next section.

4. Aggregation procedure

The aggregation clusters procedure we have implemented is based on a hierarchical scheme, as in Fee et al. (1996a). We make use of the criterion of the algorithm CLUSTER (Dubes and Jain, 1976) to evaluate the connection strength between partition couples. The results of the MEP algorithm are the conditional probabilities $P(j \mid x_k)$ and a set of centroids $\hat{Y}$. To obtain the initial partition for the aggregation procedure, we terminate the annealing by forcing β infinity, where the method becomes regular k-means. The square-error for cluster C_j, using Eq. (3), is defined as:

$$e_j^2 = \sum_{i=1}^{n_j} E(\hat{Y}_j, x_i^{(j)}), \tag{9}$$

where $x_i^{(j)}$ is the ith spike waveform belonging to cluster C_j, and n_j indicates the total number of waveforms.

The first step to reduce the excess of differentiation is to gather clusters whose centroids are near each other; the second step is to discard any clusters made up by only a few elements, reassigning the elements to the surviving clusters. These two steps form the COLLAPSER procedure. Experimentally we found that the COLLAPSER improved the performance of the aggregation procedure. After collapsing the centroids, we used the criterion of CLUSTER (see Jain and Dubes, 1988) in the aggregation procedure. Such a criterion allows the evaluation of the connection strength between couples of partitions that are consecutively joined, two at time, in order to discover a cluster solution.

4.1. COLLAPSER

COLLAPSER first inspects the size r of the obtained partition P, then looks for those clusters whose prototype vectors are 'too close' (according to a threshold value δ), or 'too weak' (the size of their clusters is too small compared with the feature space dimension).

The δ parameter is up to avoid problems related to machine precision or the rough annealing scheme used, typical values are in the $[10^{-1},10^{-4}]$ range.

The ALGORITHM of COLLAPSER is:

(i) If the cardinality r of the partition P is less than Γ, skip to the MEP algorithm with a new annealing scheme.

(ii) For all pairs of centroids $\hat{y}_{j1}$ and $\hat{y}_{j2}$ merge their corresponding clusters if $E(\hat{y}_{j1}, \hat{y}_{j2}) < \delta$ (Eq. (3)), for a suitable positive value of δ, and let $\hat{y}_{j3}$ be its cluster center, where j_3 is such that: $C_{j3} = C_{j1} \cup C_{j2}$.

(iii) For all prototypes delete $\hat{y}_j$ if its cardinality m is less than the cardinality n_j of its corresponding cluster C_j, then redistribute the elements of C_j according to the minimal distance from all prototypes.

4.2. Aggregation

Let P be the reduced partition induced on the set Ω of spikes after the COLLAPSER algorithm, $P = \{C_1, C_2,\ldots,C_r\}$, where each $C_j{:}j = [1, 2,\ldots,r]$ is a cluster having n_j elements, satisfying:

- $C_{j1} \cap C_{j2} = \emptyset$ for j_1 and j_2 from 1 to r, $j_1 \neq j_2$
- $C_{j1} \cup C_{j2} \cup \ldots \cup C_{jr} = P$.

The cardinality r of P is such that $\Gamma < r < C$, where C is the overly number of class we set in the MEP step, while Γ is the number of clusters related to the recording procedure.

A partition P_2 is nested into a partition P_1 if P_2 is formed by merging components of P_1. In the following we consider a sequence of nested clusterings where the number of clusters decreases (one at each step) as the sequence progresses. The process is iterated until the fixed number Γ is reached.

To perform the single aggregation step, a similarity measure used in the CLUSTER algorithm has been considered:

$$M_{\text{ave}} = \frac{1}{r} \sum_{j=1}^{r} \left[\left(\frac{n_j}{e_j^2} \right)^{1/2} (N - n_j)^{-1} \sum_{j1=1, j1 \neq j}^{r} n_{j1} E(\hat{Y}_{j1}, \hat{Y}_j) \right], \tag{10}$$

that is, the ratio of the 'average' distance from cluster j to all other clusters is divided by the square error for cluster j and averaged over all clusters. This measure is

a heuristic number such that the larger this number, the better the clustering. The measure seems appropriate in relation to the number of available waveforms.

The resulting ALGORITHM of AGGREGATION is:

(i) For all pairs of clusters (j_1, j_2) in the current clustering, let $P(j_1, j_2)$ be the partition resulting from the merging of clusters j_1 and j_2 into cluster j, and let $\hat{y}\ j$ be its cluster center.

(ii) Compute M_{ave} for such a partition.

(iii) Find the pair of clusters $(\hat{j_1}, \hat{j_2})$ such that:

$$M_{ave}(\hat{j_1}, \hat{j_2}) = \max(M_{ave}(j_1, j_2)),$$

where the maximum is over all pairs of clusters in the current clustering.

(iv) Merge clusters $\hat{j_1}$ and $\hat{j_2}$ to form the next clustering; if the number of clusters in the current clustering is Γ, stop, if not return to the initial step (i).

5. Experimental results

In this section we illustrate the effectiveness of the proposed unsupervised approach to clustering. The synthetic data set I–Λ (Table 1) described in (Fukunaga, 1990) was used to test our clustering technique; we present an example involving real data of multiple unit recordings from spinal neurons of gas anaesthetized rats. The entire clustering procedure was carried out in the original dimensional space of the sampled time points ($m = 8$ for I–Λ and $m = 16$ for our data).

For the first experiment, data are drawn from two multivariate ($m = 8$) Gaussian distributions having different expected vectors and covariance matrices. Both the covariance matrices are diagonal.

For this data set the Bayes error is 1.9%, while the nearest mean reclassification algorithm (Hall and Ball, 1965; Fukunaga and Koontz, 1970) was 9.5%.

One hundred samples were generated from each class of I–Λ and mixed together to form 200 samples; our clustering procedure was then applied to classify these samples in two clusters ($\Gamma = 2$); ten training sets were generated, and the experimental results averaged and the standard deviation computed.

The following setting was used:

MEP:

- $C = 4$, as the overly number of clusters.

Table 1
Data I–Λ

i	1	2	3	4	5	6	7	8
m_i	0	0	0	0	0	0	0	0
λ_i	1	1	1	1	1	1	1	1
m_i	3.86	3.10	0.84	0.84	1.64	1.08	0.26	0.01
λ_i	8.41	12.06	0.12	0.22	1.49	1.77	0.35	2.73

- $\beta_{min} = 1$; $\beta_{max} = 96$; $step = 5$; for the annealing.

COLLAPSER:

- $\delta = 10^{-2}$ threshold for the 'closeness' of prototypes.

AGGREGATION:

- no parameters required.

We obtain an error rate of 5.9% with a standard deviation of 1.3.

In the example on real data $\Gamma = 4$ was chosen as the expected number of clusters. In our data, we found that for the end points of the temperature, comprised in arbitrary units of 345 and 5°, a freezing session to be exploited in steps of 20 units could provide a satisfactory parameter value setting for the data.

The following setting has been used:

MEP:

- $C = 12$, as the overly number of clusters.
- $\beta_{min} = 5$; $\beta_{max} = 345$; $step = 20$; for the annealing.

COLLAPSER:

- $\delta = 10^{-2}$, as threshold for the 'closeness' of prototypes.

AGGREGATION:

- no parameters required.

Fig. 1 illustrates the clustering results on a spinal recording source, a projection to highlight the clusters has been selected. To properly see the anisotropy of each cluster the spectrum of variability associated with the principal components of the waveform-pair should be taken into account as thoroughly discussed in Fee et al. (1996a). The '+' signs indicate prototypes. Fig. 1a shows the scatter plot of a set of 1112 acquired waveforms. Fig. 1b shows the initial prototypes. Fig. 1c and Fig. 1d shows the prototypes after performing, respectively, 10 and 18 iterations of the MEP algorithm. Fig. 1e shows the results of the COLLAPSER, while in Fig. 1f the final result of the aggregation procedure is shown.

In Fig. 2 the waveforms corresponding to the final results of Fig. 1 are shown linked with some specific index of neuronal dynamics (Raster dot, ISI Histograms).

In the raster dot the time of occurrence of each spike is represented by a single dot. Incidentally it is possible to note alternate activations of the first two neurons that were placed at the maximum 100-μm distance, possible in mutual inhibitory connections. The ISI index is a measure of the minimal time required before a second spike firing. This parameter outlines some dynamic features of the neuron, like the membrane refractoriness time, and establishes an upper limit to the frequency of discharge, this limit being reliably obtainable from a single unit.

The clustering algorithm was run on a Sun 4/80 (SPARC station 20) running MatLab language (Math Work Inc.). The CPU mean time required to process the data of Fig. 1 was 23.6 s for the MEP, 0.18 s for the COLLAPSER and 2.7 s for the AGGREGATION.

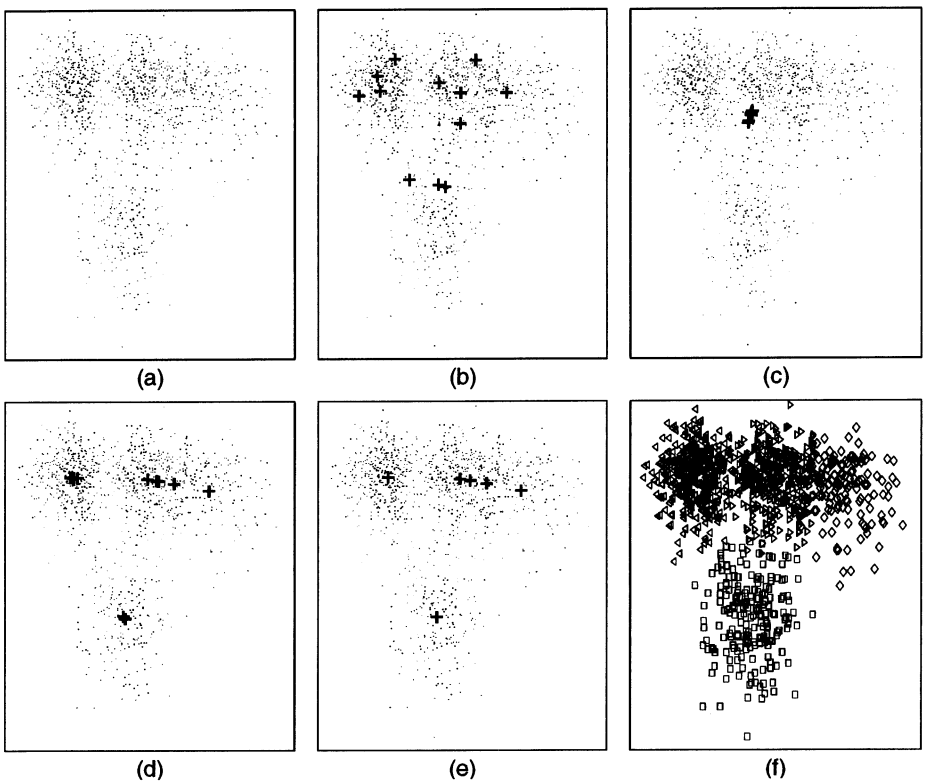

Fig. 1. Intermediate results of the clustering algorithm on a spinal cord source. (a) Original image (projection to highlight clusters), (b) initial prototypes, (c) 10 iterations, (d) 18 iterations (convergence), (e) prototype after the COLLAPSER step, (f) final result after the aggregation procedure.

6. Discussion

In this paper we have described a structure of algorithms for the unsupervised classification of neuronal signals. Three algorithms have been used in sequence and implemented in an automated routine that allows the correction of small fluctuations in data assignments. A clustering method based on the MEP has been used. The basis of our procedure is deterministic annealing, where good results can be achieved in the partitioning of the subset of elements from an original global aggregate. The initial partition of the sampled signals is obtained by a random process, in that the process is independent of the initial choice of configuration. The end points and the steps we have found seem reliable for neuronal recordings from many sources (e.g. thalamus and cortex, unpublished data) and probably represent a suitable tuning for extracellular neuronal recordings. As has been shown in applications on spin glasses, cooling, at temperatures where large clusters form, represents a limit overcome only by very slow further temperature drops (Kirkpatrick et al., 1983). Thus the process in our case is controlled until a pre-completion level. An inflated estimation or excess of clusters (three times the expected as a mean) is established so far (cluster overly). The fitting stages on the data set obtained by the incomplete annealing are forced by a hierarchical fitting method that acts by merging existing structures along a criterion of similarity between couples of partitions. The hierarchization is programmed so as to avoid expensive procedures at the final stages of cooling. This locking-in to the previous results from the MEP algorithm, generally considered a drawback, can be advantageous in our experimental schedule, due to the stability of the proximity criteria during the cluster reductive phase. The excess of differentiation, caused by incomplete annealing, has been handled by two steps: in the first, aggregating the small and improperly split clusters lets highly significant sets be analyzed by the second step; in the second step we use a function of similarity to evaluate the distances between distributions to be aggregated in a hierarchical fashion. The measure introduced by (Dubes and Jain, 1976), in their CLUSTER algorithm, which belongs to the square error clustering methods, is used to evaluate the 'best' clustering solution. This is an iterative clustering algorithm that starts with an initial partition and then reduces square error by assigning a pattern to clusters. In our recordings, vectors of 12–16 timepoints were obtained. No feature extraction was made in that no a priori knowledge was available. With this in mind we chose not to use general feature extraction methods, e.g. principal component analysis, because they are not a suitable tool for a discriminant analysis.

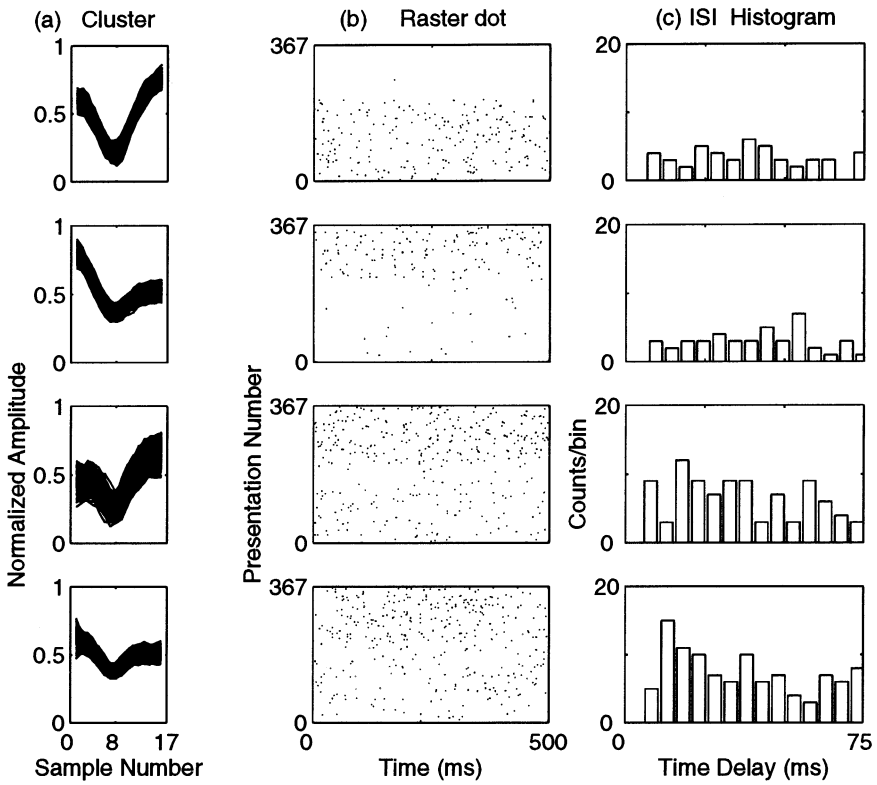

Fig. 2. (a) Representative spike waveforms for the final clusters identified in the data set and shown in Fig. 1e, (b) corresponding raster dots, (c) corresponding ISI histogram. The bin size is 1 ms.

Thus, the method presented here seems to be fast and robust and is a low-cost system to extract, with accuracy, families of spikes from a multidimensional electrophysiological recording.

Acknowledgements

We thank Dr M. Artuso for useful discussions, and B. Carey for the English revision.

References

Bezdek JC. Pattern Recognition with Fuzzy Objective Function Algorithms. New York: Plenum Press, 1981:63–70.

Biella G, Riva L, Sotgiu ML. Interaction between neurons in different laminae of the dorsal horn of the spinal cord. A correlation study in normal and neuropathic rats. Eur J Neurosci 1997;9:1017–25.

Dubes RC, Jain AK. Clustering techniques: The user's dilemma. Pattern Recognition 1976;8:247–60.

Duda RO, Hart PK. Pattern Classification and Scene Analysis. New York: Wiley, 1973:188–202.

Fee MS, Mitra PP, Kleinfeld D. Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. J Neurosci Methods 1996a;69:175–88.

Fee MS, Mitra PP, Kleinfeld D. Variability of extracellular spike waveforms of cortical neurons. J Neurosci Methods 1996b;76:3823–33.

Fukunaga K. Introduction to Statistical Pattern Recognition, 2nd ed. San Diego, CA: Academic Press, 1990:46.

Fukunaga K, Koontz WLG. A criterion and an algorithm for grouping data. IEEE Trans Computers 1970;C-19:917–23.

Hall DJ, Ball GB. ISODATA: A Novel Method of Data Analysis and Pattern Classification. Tecnical Report, Stanford Research Institute, Menlo Park, CA, 1965.

Hofmann T, Buhumann JM. Pairwise data clustering by deterministic annealing. IEEE Trans Pattern Anal Machine Intell 1997;19(1):1–14.

Jain AK, Dubes RC. Algorithms for Clustering Data. Englewood Cliffs, NJ: Prentice-Hall, 1988:108.

Jaynes ET. Information theory and statistical mechanics. Phys Rev 1957a;106:620–30.

Jaynes ET. Information theory and statistical mechanics II. Phys Rev 1957b;108:171–90.

Kirkpatrick S, Gellat CD Jr., Vecchi MP. Optimization by simulated annealing. Science 1983;220:671–80.

Masulli F, Bonguś P, Schenone A, Artuso M. Fuzzy clustering methods for the segmentation of multivariate medical images. Proc 7th Int Fuzzy Systems Assoc World Congress IFSA (Prague) 1997;3:123–8.

Mitra PP, Fee MS, Kleinfeld D. The variability of extracellular spike waveform is not random: A new algorithm for spike sorting. Soc Neurosci (Abstr) 1995;21:115.

Parzen E. On the estimation of a probability density function and the mode. Ann Math Stat 1962;33:1065–76.

Rose K, Gurewitz E, Fox GC. A deterministic annealing approach to clustering. Pattern Recognition Lett 1990;11:589–94.

Rose K, Gurewitz E, Fox GC. Constrained clustering as an optimization method. IEEE Trans Pattern Anal. Machine Intell 1993;15(8):785–94.

Sahani M, Pezaris JS, Andersen RA. Extracellular recording from adjacent neurons: I. A maximum-likelihood solution to the spike separation problem. In: Proceedings of the Society for Neuroscience 27th Annual Meeting Soc. Neurosci. Abstr. vol. 23, Part 2, p. 1546, 1997.