ELSEVIER

Journal of Neuroscience Methods 122 (2002) 43–57

JOURNAL OF NEUROSCIENCE METHODS

www.elsevier.com/locate/jneumeth

Using noise signature to optimize spike-sorting and to assess neuronal classification quality

Christophe Pouzat *, Ofer Mazor, Gilles Laurent

California Institute of Technology, Division of Biology, 139-74, Pasadena, CA 91125, USA

Received 17 May 2002; received in revised form 27 August 2002; accepted 28 August 2002

Abstract

We have developed a simple and expandable procedure for classification and validation of extracellular data based on a probabilistic model of data generation. This approach relies on an empirical characterization of the recording noise. We first use this noise characterization to optimize the clustering of recorded events into putative neurons. As a second step, we use the noise model again to assess the quality of each cluster by comparing the within-cluster variability to that of the noise. This second step can be performed independently of the clustering algorithm used, and it provides the user with quantitative as well as visual tests of the quality of the classification.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Data clustering; Expectation-maximization; Antennal lobe; Locust; Sampling jitter; Multi-electrode; Tetrode

1. Introduction

Understanding brain codes will, as a prerequisite, likely require the simultaneous sampling of large populations of neurons. While many powerful imaging techniques have been developed (e.g. membrane voltage, Wu et al., 1994; intrinsic signal, Frostig et al., 1990; fMRI, Ogawa et al., 1992), extracellular recording remains the only method that provides both single neuron and single action potential resolution from large and distributed samples. Multi-neuron extracellular recordings, however, are useful only if the spikes generated by different neurons can be sorted and classified correctly. Although a given neuron may generate spikes with unique extracellular signal features, making the identification issue trivial, in most cases, the electrophysiologist must, from noisy and ambiguous primary data, answer the following questions:

1) What is the waveform generated by each neuron, or *unit*, on each recording site?
2) How many units were sampled by the recording?
3) On what objective basis should an individual event, or spike, be classified as originating from one of the units sampled?
4) How should superpositions, due to the nearly simultaneous occurrence of two (or more) spikes, be resolved?
5) How likely are misclassifications, that is, how often is an event generated by neuron A classified as originating from neuron B, and vice versa?
6) How can we test and quantify objectively the reliability of our classification procedure?

The first three questions have been the focus of much investigation and several methods have been proposed (reviewed by Lewicki, 1998), such as principal component analysis (Glaser and Marks, 1968), Bayesian classification (Lewicki, 1994), clustering based on the Expectation-Maximization algorithm (Sahani, 1999), template matching (Millecchia and McIntyre, 1978), wavelet transform based methods (Letelier and Weber, 2000; Hulata et al., 2000) and clustering methods that use spike time information to determine cluster boundaries (e.g. Fee et al., 1996b). Question 4 has been directly

* Corresponding author. Present address: Laboratoire de Physiologie Cérébrale, CNRS UMR 8118, 45, rue des Saints Pères, 75006 Paris, France. Tel.: +33-1-4286-3831; fax: +33-1-4286-3830

E-mail address: christophe.pouzat@biomedicale.univ-paris5.fr (C. Pouzat).

0165-0270/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0165-0270(02)00276-5

addressed in two studies (Atiya, 1992; Lewicki, 1994). The reliability of some of these spike sorting procedures has also recently been tested empirically, using simultaneous extra and intracellular recordings (Wehr et al., 1999; Sahani, 1999; Harris et al., 2000). These later studies fail to address the main concern of the present paper: how can one evaluate, from the extracellular data *alone*, the reliability of the sorting procedure? The potential causes of unreliable spike-sorting are numerous; several are described in detail by Lewicki (1998). According to Lewicki (1998, p. 74), “Many algorithms work very well in the best case, when most assumptions are valid, but can fail badly in other cases. Unfortunately, it can be difficult to tell which circumstance one is in”. The simple tests we present here are an attempt to address this dilemma.

In the body of the paper, we will provide a detailed description of our methods, as well as an illustration of their use on in vivo recordings from locust antennal lobe neurons. We begin by presenting a brief description of the experimental procedure including data collection. Next, we describe the method for generating a model of the experimental noise and for testing the accuracy of the model. We then proceed to show how that model can be used first to cluster spikes, and then to test the quality of the classification. Finally, we run the entire procedure on an example of real data.

2. Methods

2.1. Data collection and representation

2.1.1. Preparation and recordings

All experiments were carried out on adult locusts (*Schistocerca americana*) of both sexes, taken from a crowded colony and prepared as described in Laurent and Naraghi (1994).

Neurons were recorded with silicon probes from the Center for Neural Communication Technology of the University of Michigan (Drake et al., 1988). A diagram of the probe tips with 16 recording sites is shown on Fig. 1A. The probe was connected to a custom made impedance matching preamplifier. The preamplifier was connected to two 4 channels differential AC amplifiers (AM model 1700 AM Systems, Carlsborg, WA). The signals were bandpass filtered between 300 and 6000 Hz and amplified 10 000 times. Data were acquired at 15 k samples/s using a 12 bit A/D card (Win30 D, United Electronics, MA).

Data with a good signal to noise (S/N) ratio were collected relatively close to the surface (50–100 μm) of the antennal lobe (AL). Spikes recorded in the AL were attributed to the activity of projection neurons (PNs), as the AL contains only two neuron populations: the PNs, which are the output neurons and fire Na^+ action potentials and the local neurons (LNs), which are axonless and fire no Na^+ action potential (Laurent and Davidowitz, 1994). We were unable to record clear spikes with the silicon probe from the antennal nerve or its projections into the AL. Afferent axons are very small and numerous (90 000), precluding clear discrimination of single neuron signal from noise.

2.1.2. Data processing

Data were analyzed offline on a PC computer (Pentium III 550 MHz, 256 MB RAM) using Igor (WaveMetrics, Lake Oswego, OR) or Scilab (a free Matlab-like software package available at: www-rocq.inria.fr/scilab). All the routines were custom developed (or are freely available on the world-wide-web, see below) and are available upon request.

2.1.3. Event detection

For the detection stage only, the traces were first digitally smoothed (3-point Box Filter). Events (i.e. putative spikes) were then detected as local maxima with a peak value exceeding a preset threshold. In cases where the spike peak occurred at slightly different times on different recording sites, only one time value was used: the time from the site with the largest peak amplitude. The detection threshold was set as a multiple of the standard deviation (S.D.) of the whole trace. We typically used thresholds between 2.25 and 3.5 S.D.s.

2.1.4. Event representation

Detected events can be represented in many different ways (Lewicki, 1998). Yet, the choice of a representation can strongly influence both the speed and the reliability of the classification procedure. In general, one measures a set of D parameters for each event; each event thus becomes a point in a D dimensional space. This space is called *event space*. Our goal was to optimally predict the effect of recording noise on the distribution of points that represent events in event space. Unfortunately, several common parameter choices, such as peak and valley amplitudes or half width are computed by differentiating the raw data. This makes signal-noise separation difficult.

We therefore chose to represent each event as follows. A sweep of d consecutive sample points around the peak of the event was examined from each recording site. For our setup, we set $d = 45$ (equivalent to 3 ms), with the peak aligned to the 15th position. The sweeps were then concatenated. Therefore if one labels the successive amplitudes of an event on site A, $A_1A_2...A_{45}$, on site B, $B_1B_2...B_{45}$, on site C, $C_1C_2...C_{45}$ and on site D, $D_1D_2...D_{45}$, the vector representing the event was:

$$\boldsymbol{e} = (A_1...A_{45}B_1...B_{45}C_1...C_{45}D_1...D_{45})^T,$$

where the superscript T means transpose. For our purposes, therefore, the dimensionality of the event

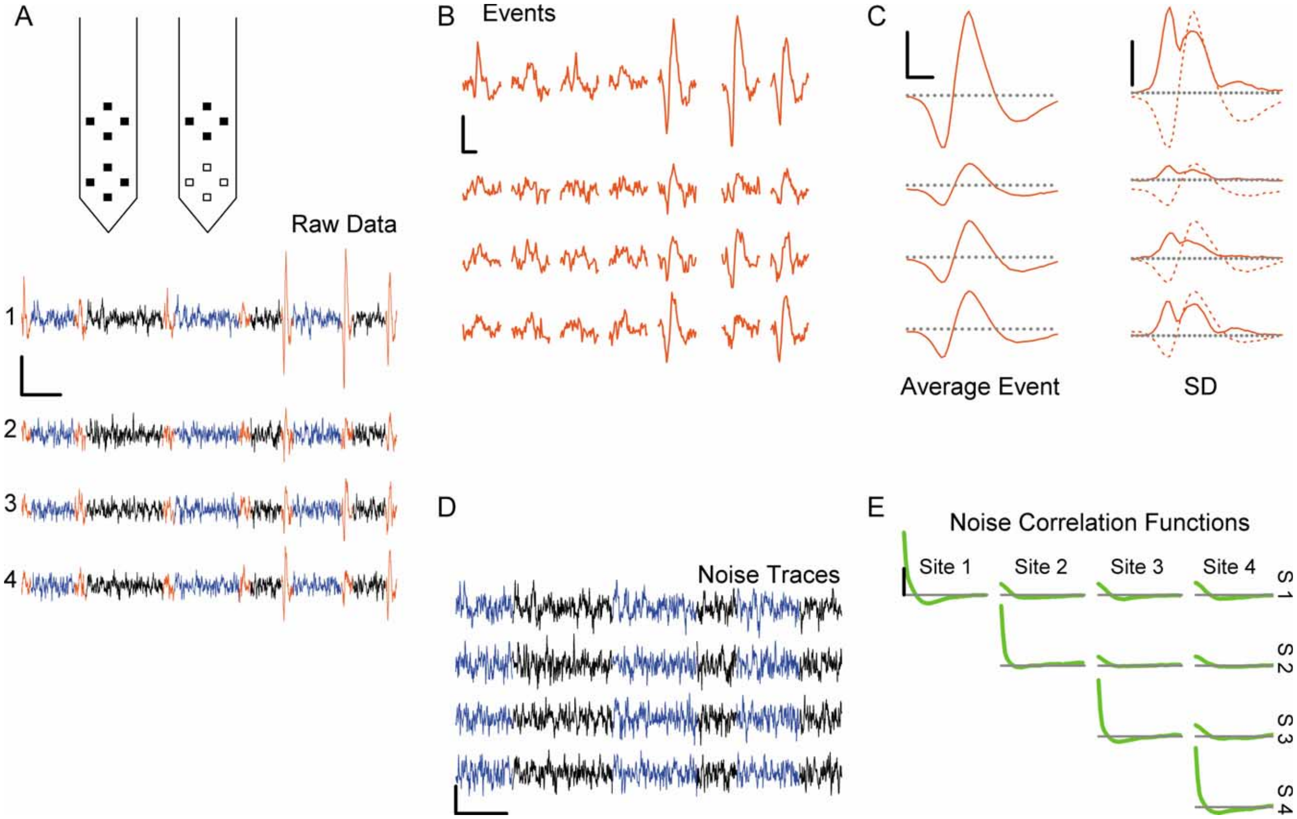

Fig. 1. A Top, scheme of the recording probe tips. Each shank carries two tetrodes. The four tetrodes are identical. The surface of each recording site (filled and open squares) is 177 μm^2. The recording sites are placed at the corners of a square with a 50 μm diagonal. The center to center distance between two neighboring tetrodes is 150 μm. The shank width is 83 μm. Bottom, 1 s of data bandpass filtered between 300 and 6000 Hz. Site 1 is the lowest site of the tetrode; the other sites are numbered counter-clockwise. Calibration: vertical, 100 μV; horizontal, 100 ms. Detected events are shown in red. The traces are displayed inverted, that is, a positive deviation on the trace corresponds to a negative deviation with respect to the reference potential. B, the seven events shown on a smaller temporal scale (vertical, 100 μV; horizontal, 1 ms). A sweep, 3 ms long, has been built with the peak of each event at 1 ms. C left, average event computed from the 1493 events recorded over 60 s. Vertical, 50 μV; horizontal, 0.5 ms; dotted line, 0 μV. C right, corresponding standard deviation. Vertical, 25 μV; same time scale as C left. Dotted line: S.D. level expected from the noise (15.4 μV on site 1, 15.1 μV on site 2, 14.6 μV on site 3 and 15.0 μV on site 4). Red dashed curves: average event (same as C left) on the same site for comparison. D, construction of the noise traces. After removal of 3 ms of data around each event, the remaining data are concatenated. Vertical, 50 μV; horizontal, 100 ms. E, Noise correlation functions computed from the noise traces. Upper diagonal matrix display: auto-correlation functions on diagonal and cross-correlation functions otherwise; 3 ms sweep. Vertical, 0.01 μV^2.

space, D, is 180 (4×45). It will become clear that with this choice of event space, the effect of noise on the distribution of events can be easily predicted. Note that our initial peak detection for event selection introduces some sampling-induced jitter. We will ignore this for now and show later how it can be canceled.

2.1.5. *General model for data generation*

Following Lewicki (1994) and Sahani (1999) we use an explicit model for data generation. The general assumptions in our model are:

1) The spike waveforms generated by a given neuron are constant.
2) The signal (i.e. the events) and the noise are statistically independent.
3) The signal and noise sum linearly.
4) The noise is well described by its covariance matrix.

Assumption 1 is a working approximation, appropriate for some documented cases (Fee et al., 1996a, Fig. 2; Harris et al., 2000, Fig. 4). It also applies to our recording conditions (see results). Assumptions 2 and 3, stated for completeness, are implicit in most already available spike-sorting methods and mean that the amplitude distribution of the recorded events can be viewed as the convolution of a ‘pure’ signal with the noise distribution. We can restate our hypothesis as follows: in a noise free recording, all events generated by one unit would give rise to the *same* point in event space. In a noisy recording, however, events generated by one unit would give rise to *a cloud* of points centered on a position representing the ‘ideal’ waveform of the unit. The distribution of the points should be a multivariate Gaussian whose covariance matrix would be the noise covariance matrix *regardless* of the position of the unit in event space.

2.2. Noise model

2.2.1. Noise covariance matrix

To measure the statistical properties of the noise, we began by removing from the raw traces all the detected events (i.e. all the d-point sweeps) and concatenating all the inter-event traces. We call the resulting waveforms 'noise traces' (see Fig. 1D). The auto-correlation function was then calculated for each recording site (diagonal, Fig. 1E), as were the cross-correlation functions between all pairs of sites (Fig. 1E). These correlations were only computed within continuous stretches of noise (i.e. the discontinuities in the noise traces due to eliminated spikes were skipped). In addition to recording noise, these cross-correlations will also account for any cross-talk between recording channels (Zhu et al., 2002).

In event space the auto- and cross-correlation functions translate into the noise covariance matrix which was build by blocks from these functions as follows (we refer here to the four recording sites as site A, B, C and D):

$$\begin{pmatrix} AA & AB & AC & AD \\ BA & BB & BC & BD \\ CA & CB & CC & CD \\ DA & DB & DC & DD \end{pmatrix} = \Gamma,$$

where each block is a symmetric Toeplitz matrix built from the 'corresponding' correlation function (e.g. AA is a 45×45 matrix whose first row is the noise autocorrelation function on site A, AB is a 45×45 matrix whose first row is the noise cross-correlation function between sites A and B, etc.). BA is symmetrical to AB.[1]

2.2.2. Noise whitening

In order to simplify calculations and reduce the computational complexity of our algorithm, we chose to make a linear transformation on our event space (and therefore on all the detected events). The transformation matrix, U, is chosen specifically so that after transformation, the variance due to noise will be uncorrelated across dimensions (i.e. the noise covariance matrix will be the identity matrix, I). Mathematically, U has the property that

$$\Gamma^{-1} = U^T U, \qquad (1)$$

where Γ is the noise covariance matrix in the original event space. A transformation matrix, U, with this property will always exist as long as the covariance matrix (Γ) is symmetric and positive definite (which it is by definition). The matrix U is obtained from Γ^{-1} with a Cholesky decomposition (Brandt, 1999, pp. 479–484). A critical feature of the noise-whitened event space is that if our assumption (4) is correct (that the noise is well described by its second-order statistics), then the variance due to noise will be the same in every dimension with no correlations across dimensions (i.e. the cloud due to noise should be a hypersphere).

2.2.3. Test of noise model

To test assumption (4), we generated a large sample of d-point long events from the noise traces. These 'noise events' were taken from a different portion of the noise traces than was used to calculate the noise covariance matrix. Since these events should contain all 'noise' and no 'signal' (i.e. no spikes), these points will form a cloud around the origin in the noise-whitened event space and the distribution of these points around the origin will be fully described by the true statistics of the recording noise. We can now test if the second-order noise statistics (the covariance matrix) are sufficient in describing the actual noise distribution. We do this by computing the distribution of Mahalanobis distances (just the euclidean distance squared in noise-whitened space), between each noise event and the origin. In a white, Gaussian distribution, the distribution of Mahalanobis values will be a χ^2 distribution with D degrees of freedom. For our data, as we will describe in Section 3, this is indeed the case.

Testing the second-order statistics is not a guarantee that the noise distribution does not have significant higher-order moments. To check for this possibility, we measured the third momentum distribution from another pool of whitened noise events. We randomly selected 500 (or more) triplets of coordinates among the 180^3 possible ones (for an event space of 180 dimensions). If we write $\boldsymbol{n}_i = (n_{i,1}, \ldots, n_{i,180})^T$ for the ith sweep of the noise sample and if, for example, (28, 76, 129) is one of the triplets, the third moment for that triplet is obtained as follows (assuming a noise sample of size 2000):

$$\frac{1}{2000} \sum_{i=1}^{2000} n_{i,28} \cdot n_{i,76} \cdot n_{i,129}. \qquad (2)$$

We will show in Section 3 that for our data, this statistic was not significantly different from zero.

[1] For readers unfamiliar with Toeplitz matrices, we illustrate the concept using the simple case where there are only three sample points per sweep. If the auto-correlation function on site A is the vector $(\alpha \quad \beta \quad \gamma)$, then AA would be

$$AA = \begin{pmatrix} \alpha & \beta & \gamma \\ \beta & \alpha & \beta \\ \gamma & \beta & \alpha \end{pmatrix},$$

that is $AA_{i+1,j+1} = AA_{i,j}$, for $i \geq j$.

2.3. Noise model-based clustering

2.3.1. Specific data generation model

If our first assumption about data generation is correct (that spike waveforms are constant), the distribution of events in event space, after the linear coordinate transformation (Eq. (1)), should be a set of clouds of identical shapes (hyperspheres), each centered on its underlying unit.[2] Our goal is now to determine the number of such clouds and the position of their centers in event space.

To this end, we introduce a *specific data generation model* (M) that extends the general data generation model by specifying the number of units, K, their waveforms and their discharge frequencies. In event space, the waveforms of the K units translate into a set of K vectors $\boldsymbol{u}_j$ (joining the origin to the point representing unit j, $j \in \{1, \ldots, K\}$). Our goal is to find the model that gives the best explanation of the data sample $S = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N\}$. A common and efficient way to do this is to find the model which maximizes a posteriori the probability to observe the sample actually observed, i.e. to maximize the likelihood function (Brandt, 1999; Bishop, 1995).

The likelihood function is computed under our assumptions and in the noise-whitened coordinate system as follows. We first compute the probability (density) for unit $\boldsymbol{u}_j$ to have generated event $\boldsymbol{e}_i$, $p(\boldsymbol{e}_i|\boldsymbol{u}_j)$. For that we introduce the residual vector Δ_{ij}:

$$\Delta_{ij} = \boldsymbol{e}_i - \boldsymbol{u}_j, \tag{3}$$

then

$$p(\boldsymbol{e}_i|\boldsymbol{u}_j) = \frac{1}{(2\pi)^{\frac{D}{2}}} \cdot \exp\left(-\frac{1}{2} \cdot \Delta_{ij}^T \Delta_{ij}\right). \tag{4}$$

The probability P_i for the model to have generated event $\boldsymbol{e}_i$ can now be written as a weighted sum of terms like (4), one for each of the K units of the model:

$$P_i = \sum_{j=1}^{K} \pi_j \cdot p(\boldsymbol{e}_i|\boldsymbol{u}_j), \tag{5}$$

where π_j is the probability for unit j to occur, i.e. the ratio of the number of events from unit j to the total number of events in the sample, N. The a posteriori probability to observe the sample is then, assuming independence of the N sample elements, the product of the probabilities to observe each one of them separately:

$$\mathscr{P}(S;\ M) = \prod_{i=1}^{N} P_i. \tag{6}$$

The likelihood function is simply the logarithm of $\mathscr{P}$:

$$\mathscr{L}(S;\ M) = \sum_{i=1}^{N} \ln(P_i). \tag{7}$$

Several iterative algorithms exist to maximize $\mathscr{L}$ (Redner and Walker, 1984; McLachlan and Krishnan, 1997). We used the Expectation-Maximization algorithm (EM algorithm, formalized by Dempster et al., 1977, and introduced in the electrophysiological literature by Ling and Tolhurst, 1983). The EM algorithm is very simple in the present context, fairly intuitive (Bishop, 1995) and its convergence to local maxima has been proven for the present model (without outliers in the sample: Boyles, 1983; Wu, 1983). Moreover, for our typical data samples, outliers do not appreciably affect the speed or accuracy of the algorithm.

The standard EM algorithm finds the 'best' model for a given number of units. It does not provide, by itself, the actual number of units, K, present in the data sample. Several criteria have been proposed in the statistical literature to perform this task (for an overview, see Fraley and Raftery, 1998 (especially Section 2.4), and Biernacki and Govaert, 1998). Among the methods we tried, however, we found that the *Bayesian Information Criterion* (BIC), proposed by Schwarz (1978), worked well for our data (where most clusters are well separated in event space). The BIC penalizes an increase in the number of components by subtracting from $\mathscr{L}$ a term proportional to $v \cdot \ln(N)$, where N is the number of sample events and v is the number of model parameters. We then simply keep the model with the value of K which maximizes the BIC[3] (Fraley and Raftery, 1998).

2.3.2. Event classification

Once a model is established, we attribute each event, $\boldsymbol{e}_i$ to one of the K units, by finding the j that minimizes $|\Delta_{ij}|^2$. The rationale is the following: if unit $\boldsymbol{u}_j$ has indeed generated event $\boldsymbol{e}_i$ then the components of the residual vector Δ_{ij} are random numbers drawn from a multivariate Gaussian distribution and the probability to observe $|\Delta_{ij}|^2 = \Delta_{ij}^T \cdot \Delta_{ij}$ is given by a χ^2 distribution with D degrees of freedom (assuming noise whitening has been performed). By choosing the unit producing the

[2] We should expect some outliers as well, due to nearly coincident spikes.

[3] A free software package is available that includes an implementation of the EM algorithm with the BIC. 'Mixmod', written in C++ by Biernacki, Celeux, Govaert, Langrognet and Vernaz is available at the following address: www.inrialpes.fr/is2/software/MIXMOD/.

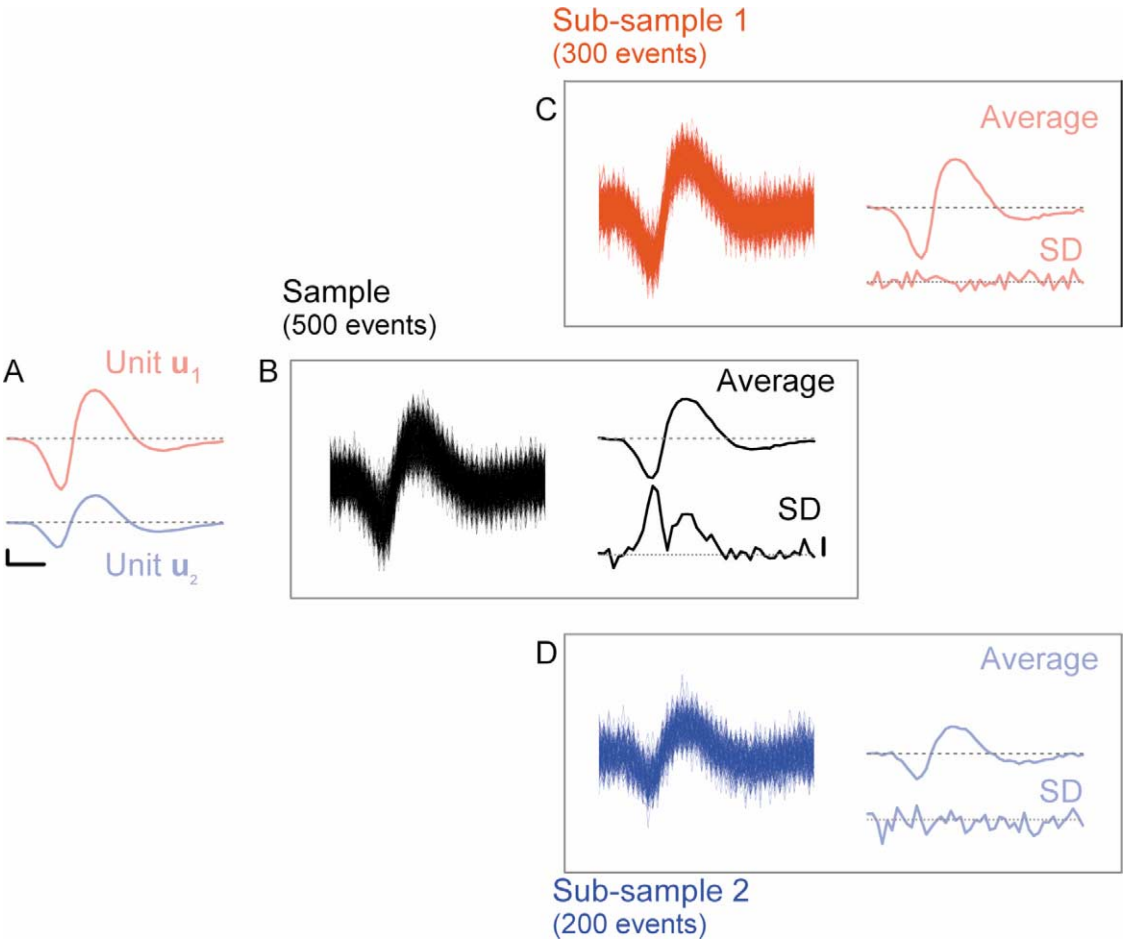

Fig. 2. Illustration of the S.D. test on simulated data. A, waveforms of the two units used to generate the data (see text). The scale bars are arbitrary. The vertical bar has the value 1, equal to the noise's S.D. To compare with real data, the length of the horizontal bar would be 0.5 ms; dashed line at zero. B left, 500 events generated from the two units (300 from $\boldsymbol{u}_1$ and 200 from $\boldsymbol{u}_2$) by adding Normal white noise to the units waveforms. B right top, average event computed from the 500 events of the sample. B right bottom, SD computed from the sample. Dotted line, 1 (expectation from the noise properties). Vertical scale bar, 0.1. Notice the non-zero SD at the peak and valley of the average event. C and D, as in B, except that C and D have been built from the sub-samples generated by unit $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, respectively. Note the reduction of S.D. variations.

smallest $|\Delta_{ij}|^2$ we are simply choosing the unit with the highest probability to have generated the event.[4]

For some events, even the smallest $|\Delta_{ij}|^2$ was very unlikely given the χ^2 distribution (e.g. in the 99th percentile). In these cases, we looked for the superposition of any two units, e.g. $\boldsymbol{u}_j$ and $\boldsymbol{u}_l$, which gave the smallest $|\Delta_{i,j+l}|^2$ value. To this end we tested all possible pairs of units and all relative timings of their peaks. This was easily computed for we knew the entire waveform of each unit. This approach was formalized by Atiya (1992) and an alternative method to resolve superpositions has been proposed by Lewicki (1994). If, after this step, we still did not find a small enough $|\Delta_{i,j+l}|^2$, we classified the event as an outlier.

2.4. Model verification tests

No matter how much effort is devoted to optimizing model-generation and event-classification procedures, in the end it is always possible for the results of a spike-sorting routine to be sub-optimal. In many recordings there may be pairs of neurons whose spike waveforms are close enough (with respect to the size of the noise) that their events could never be accurately distinguished. In such a case, some algorithms may lump the pair into one cluster and others might split such a pair in two. In either case, an experimenter would like to detect such a situation, and if the pair of units is really inseparable, discard the spikes from those cells or treat them as multi-unit data. Furthermore, due to the complexity of the task, even the best algorithms will occasionally generate incorrect models when given reasonable data. Again, this is a situation one would like to detect.

For this reason, we developed three tests for assessing the quality of spike-sorting results on a cluster-by-

[4] Strictly speaking, we should choose the unit giving the largest product $\pi_j \cdot p(\boldsymbol{e}_i|\boldsymbol{u}_j)$, but it turns out that our units are typically far apart. Therefore a unit $\boldsymbol{u}_i$ which does not generate $\mathbf{e}_i$ give a $|\Delta_{il}|^2$ value much too large, which is equivalent to a negligible $\pi_i \cdot p(\boldsymbol{e}_i|\boldsymbol{u}_i)$ value.

cluster basis. Since we have a quantitative model of data generation, we can use it to make detailed predictions about the properties of our classified data. Here we illustrate the tests' principles by applying them to simulated data. In Section 3 we will present the same test applied to real data.

Consider the simple situation in which we record from a single site and where only two units, $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, are active. Assume also that both units fire at low rates, so that nearly simultaneous spikes from unit $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are rare. The original waveforms of the two units (used to generate the data) are shown in Fig. 2A. During our 'recording session', we sample 500 events, superimposed in Fig. 2B (left). Each event corresponds to one of the units, to which random 'noise' drawn from a Normal distribution has been added to each of the 45 sampling points. This artificial data generation procedure is such that our model assumptions apply exactly to the sample (in this case, the noise is already white). In this sample, 300 events have been generated from unit $\boldsymbol{u}_1$ and 200 from unit $\boldsymbol{u}_2$.

For the first two tests, we will consider two potential models of data generation. In the first case, all events of the sample are (incorrectly) classified as coming from a single unit; in the second case, the data generation model contains the two units $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, and all events are correctly classified.

2.4.1. The S.D. test

The mean event and the S.D. computed from all 500 events of the sample are shown on the right of Fig. 2B. Note how the S.D. varies, reaching maxima at times where the two waveforms ($\boldsymbol{u}_1$ and $\boldsymbol{u}_2$) differ the most. Based on our initial assumptions, we would predict that this 'cluster' of 500 events could not all come from the same unit. If this were the case, all the spike-to-spike variance would be due to noise, which should be constant throughout the time course of the spike.

If we now split the sample into two correctly classified sub-samples, one consisting of the 300 events generated by unit $\boldsymbol{u}_1$ and the other from the 200 events generated by unit $\boldsymbol{u}_2$, the SD computed on the corresponding subsamples is now 'flat', centered on the background noise level (Fig. 2C,D). This matches precisely with what our model predicts for correctly classified clusters: all the spike-to-spike variability is due entirely to noise.

In this way, we can use this as a qualitative test of both the accuracy of the model and a proper classification of the events.[5] After the events have been classified, the S.D. of every cluster can be tested. Any cluster whose S.D. values deviate significantly from the S.D. of the noise can be eliminated from further analysis (or at least scrutinized more closely). In our experience, this test is quite sensitive and can routinely detect clusters that contains multiple units, even if those units are not well-separable (see Section 2.4.3).

As a final note, this test will also reliably indicate if a significant number of spikes from a small unit were not detected. This situation can arise when the peak voltage of a unit's waveform is just at the spike detection threshold. In such a case, a significant percentage of that unit's events will not be detected due to noise fluctuations. The spikes from this unit that are detected will have positive noise values near the peak, and therefore less noise variability along this portion of the waveform. This situation is therefore characterized by a 'dip' in the S.D. near the peak of the waveform, and we routinely observe this effect empirically. Hence, a cluster that exhibits a constant S.D., equal to that of the noise, is consistent with a good model together with correct spike detection and classification.

2.4.2. The χ^2 test

In this test we test the prediction that each cluster of events forms a D-dimensional Gaussian distribution. For every unit, $\boldsymbol{u}_j$, we can compute the distance from it to all events, $\boldsymbol{e}_i$, that were attributed to it. If the prediction is accurate, the distribution of the squares of these distances should follow a χ^2 distribution with D degrees of freedom.

The test is illustrated in Fig. 3A. In the first case (one-unit model), we take the sample mean as an estimate of the ideal underlying unit. We illustrate the computation of the residual of event #400 with such a model (Fig. 3Ai). Because we have 500 events in the sample, we obtain 500 χ^2 values. In Fig. 3Aii we plotted the cumulative distribution of these 500 χ^2 values (continuous gray curve). This empirical distribution can be compared with the expected one (dashed black curve). In this case, the expected distribution is a χ^2 distribution with $D-1$ degrees of freedom (i.e. 44), for we have used the average computed from the same sample.

In the second case (two-unit model), we take the averages computed from the two subsamples as estimates of the underlying units (Fig. 2C,D). The classification of event #400 is illustrated in the middle part of Fig. 3Ai. In this case, the first value suggests an unlikely event (i.e. the noise would not be expected to cause such a large deviation from the underlying unit) so the event is classified as originating from unit $\boldsymbol{u}_2$. We thus obtain from the 500 events, two empirical χ^2 distributions (Fig. 3Aii), one corresponding to subsample 1 (red curve) and one corresponding to subsample 2 (blue curve). It is clear that these two empirical distributions are much closer to the expected one. A good classification (together with a good model) should thus yield K distributions, for a model with K units, centered on a single predictable χ^2 distribution. Like the S.D. test, this

[5] This test was initially proposed, in a different context, by Jack et al. (1981).

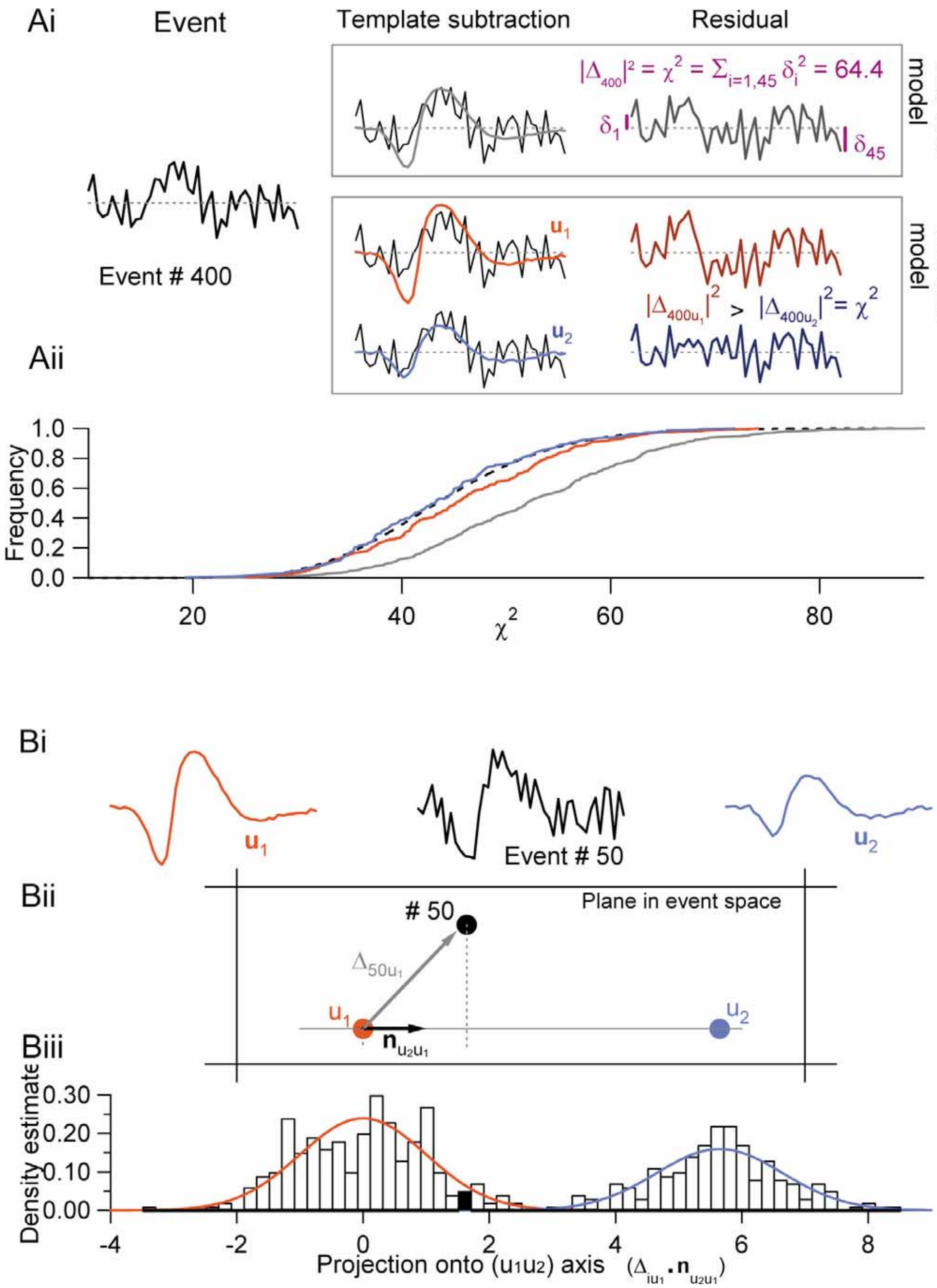

Fig. 3. A, illustration of the χ^2 test using the same computer generated sample as in Fig. 2. i, one-unit model (top): the average event is first subtracted from event #400, to yield the residual. The integral of the square of the residual is the 'χ^2' value of event #400. Two-unit model (bottom): two units could now have generated event #400; the waveforms of these units are given by templates $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ (see text). The integrals of the square of the two residuals ($|\Delta_{400\boldsymbol{u}_1}|^2$ and $|\Delta_{400\boldsymbol{u}_2}|^2$) are compared; the smallest indicates which one of the two units most likely generated the event, with its associated χ^2 value. Aii: cumulative distributions of the χ^2 values under the one- and two-unit model assumptions and their expectation (dashed line). Grey line: one-unit model ($n = 500$); red line: unit $\boldsymbol{u}_1$, two-unit model ($n = 300$); blue line: unit $\boldsymbol{u}_2$, two-unit model ($n = 200$). B, Projection test. Bi: template $\boldsymbol{u}_1$ (red), $\boldsymbol{u}_2$ (blue) and event #50 from the 500 computer generated samples. Bii: same objects in the plane that contains all three points. The straight line joining the two units has been drawn as well as the unit vector originating in u_1, $\boldsymbol{n}_{u_2u_1}$. The vector joining u_1 to event #50, Δ_{50u_1} (representation of the residual in event space) has been drawn as well. Biii: projection histogram of the 500 events of the sample. The bin containing the projection of event #50 has been filled. The red curve is probability density function (PDF) expected from the projections of the 300 events generated by unit $\boldsymbol{u}_1$ (60% of the sample) while the blue curve is the corresponding PDF expected from the projections of the 200 events generated by unit $\boldsymbol{u}_2$ (40% of the sample).

test is especially sensitive to the grouping of two similar-looking units into a single cluster and will produce a significant rightward shift in such situations.

2.4.3. *The projection test*

According to our model assumptions, the events generated by a given unit should form a cloud of points centered on the unit in event space. The precise distribution of these points should be, after noise-whitening, a multivariate Gaussian with a covariance matrix equal to the identity matrix. Moreover, the projections of two subsamples onto the axis joining the two units which generated them should form two Gaussian distributions of S.D. = 1 centered on the two units. We can now quantitatively define the distinguishability of the two units by setting a limit on the acceptable overlap between theses two distributions (overlap between distribution and event misclassifications are indeed equivalent). For instance we can decide that if more than 5% of the events coming from unit $\boldsymbol{u}_1$ or $\boldsymbol{u}_2$ are misclassified, then the two units are not distinguishable.

Fig. 3B illustrates the procedure with simulated event #50. As before, we are working in a 45 dimensional space. Fig. 3Bi shows template $\boldsymbol{u}_1$, event #50 and template $\boldsymbol{u}_2$ as voltage traces over time. Fig. 3Bii illustrates the same objects in the plane in event space that contains all three. In this plane, the straight line going through u_1 and u_2 has been drawn (u is a point in event space and $\boldsymbol{u}$ is a vector joining the origin to u) as well as the unit vector going from u_1 to u_2. We can compute the projection onto the (u_1 u_2) axis for each of the 500 events of the sample. Then according to our model assumptions the empirical distribution of the projections should be given by two Gaussian with S.D. = 1, centered on points u_1 and u_2, respectively. If two units can be reliably distinguished the two distributions will not overlap. The amount of overlap between these two Gaussian is simply a function of their Euclidian distance, making it easy to convert misclassification percentage into a minimum inter-unit distance, below which a pair of units is considered non distinguishable (e.g. assuming the same total number of events for the 2 units, a fraction of misclassification smaller than 5% requires a separation of 2.5 S.D.s between the means of the distributions; similarly, a 5 S.D.s separation would yield a 1% misclassification fraction).

A second feature of this test is that it can also detect whether a single unit has been incorrectly split into two different clusters. Consider the scenario where the spikes from $\boldsymbol{u}_1$ are split into two different clusters with centers (u_1' and u''_1). The projection between these two clusters, will form a single Gaussian distribution centered at u_1, rather than the two Gaussians predicted by the test. In this way, the projection test is most sensitive at detecting whether two clusters are inseparably close and whether a single unit has been split between two different clusters.

2.5. Sampling jitter cancellation

One final problem to solve originates from the limited sampling frequency used during data acquisition. It is obvious that the computer's clock is not synchronized with the neurons' firing: the events will be sampled with a random delay between their peak and the nearest tick of the computer's clock. While this effect may sound like a purely theoretical concern, it can have a significant effect on the tests we describe under standard recording conditions. This sampling effect and its effects on the SD and χ^2 test are illustrated in Fig. 4. We will consider this problem in the absence of recording noise, although the problem (and its solution) exists in 'real', noisy recordings (see below).

An ideal waveform from a single recording site is considered in Fig. 4A and B. The ideal waveform is made of 450 points; we show two events obtained by sampling the ideal waveform once every ten points (Fig. 4A1, only the central part of the waveform is shown). The peak of event 1 occurs at point 19 (from the origin of the sample), while the peak of event 2 occurs at point 18 (Fig. 4A1). When we build the sweeps associated with the sampled events (Fig. 1B) we align them on their peaks, causing a slight distortion, illustrated in Fig. 4A2. We see here that two sampled events arising from the same underlying waveform have different onsets and offsets. The effect of this sampling jitter on the S.D. is illustrated in Fig. 4A3. One hundred such events were generated by sampling the ideal waveform, drawing each sweep origin from a uniform, discrete, distribution between points 1 and 10 on the ideal waveform. The sweeps were aligned on their peaks and the mean event and S.D. were computed. A marked increase of the S.D. is obvious around the times when the derivative of the mean event is significantly different from 0. It is easy to show that this S.D. increase is proportional to the derivative of the underlying waveform multiplied by the sampling period. This S.D. increase caused by the sampling jitter will also result in an increase in the χ^2 obtained after template subtraction (Fig. 3A) and corrupt our model's tests. We must therefore cancel the sampling jitter.

This is done simply by using the optimal interpolation filter to recover the 'full' sweeps from the sampled sweeps, before realignment. This filter is the sinc function ($\sin(x)/x$) with a period equal to twice the sampling period (Papoulis, 1980). Fig. 4B2 illustrate this interpolation procedure and its result. The 45 sample long sweep of Fig. 4B1 has been used to build a 450 points long 'interpolated' sweep (red curve); the ideal waveform used to generate the 45 sample long sweep is shown as well (black curve) and the shift between the two curves is precisely the sampling jitter.

In practice, before running the tests we canceled jitter on every event classified as belonging to a single cluster (i.e. no outliers or superpositions). For each such event, $\boldsymbol{e}_i$, from cluster $\boldsymbol{u}_j$, we first interpolated nine points in between each true sample point to create an 'ideal' waveform. We then aligned this 'ideal' wave to its cluster

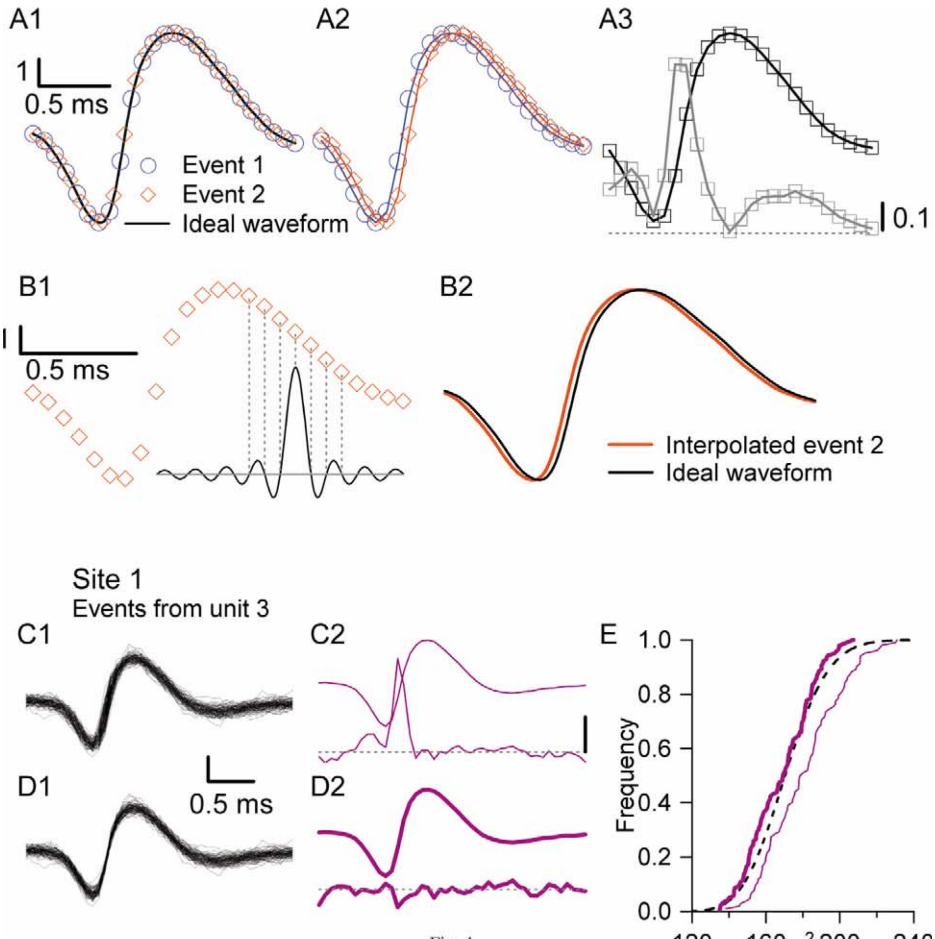

Fig. 4

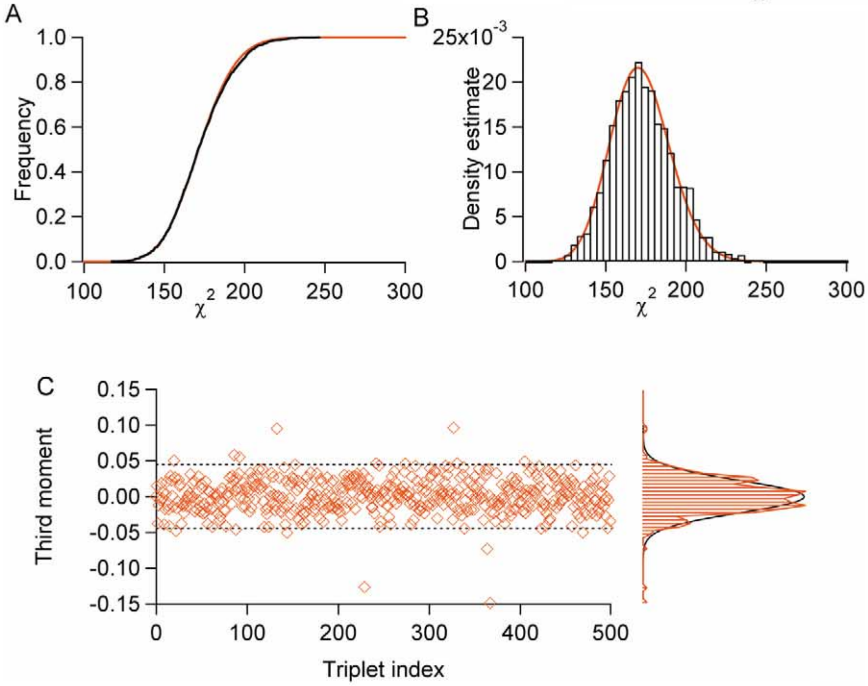

Fig. 5

Fig. 4

mean by minimizing the resulting $|\Delta_{ij}|^2$. This is illustrated with real data in Fig. 4C. Fig. 4C1 shows the events before jitter cancellation and Fig. 4C2 shows the mean event and the S.D. Fig. 4D1 shows the same events after jitter cancellation (performed simultaneously on the four recording sites) and Fig. 4D2 the corresponding mean event and S.D. The S.D. increase during the rising phase disappeared, replaced here by a slight over-fit (dip in the S.D.). Fig. 4E shows the cumulative squared distances distributions before (thin line) and after (thick line) jitter cancellation.

Finally, caution should be used when working with short sweeps, that is, sweeps whose amplitudes at the ends are not at baseline level. Performing a sinc function interpolation in such cases will generate artifactual wiggles on the interpolated sweeps. If one wants to work with such short sweeps a cubic spline interpolation is recommended.

3. Results

We now illustrate the generation and testing of a model using data recorded from the locust antennal lobe.

3.1. Data properties

3.1.1. Events

A typical, 1s long stretch of data recorded from the locust antennal lobe and band-pass filtered (see Section 2) is shown in Fig. 1A. Traces 1–4 originate from the four neighboring recording sites of one of the four tetrodes in the probe shown above. Three milliseconds around each detected event have been colored red. The remainder of each trace, shown alternately in blue and black will, from now on, be considered as noise.

The seven events detected in these four traces are shown in Fig. 1B. Each sweep is 3 ms long (45 data samples) with the peak of each event at 1 ms. Fig. 1C shows the mean (left) and corresponding S.D. (right) computed from all 1493 events so detected during 60 s of continuous data acquisition. The horizontal dotted line (right panel) indicates the S.D. expected from the background noise. Marked excesses of the S.D. are observed around the valleys or peaks of the mean event. Such excesses could have two non exclusive origins:

1) Two or more units with different spike shapes are present in the data
2) A given unit generates spikes of variable amplitudes or shapes

Causes (1) and (2) should generally depend on the cell types present as well as on the region of the neuron from which the signal is recorded. Neurons producing high frequency bursts for example, often exhibit spikes of decaying amplitude during a burst (e.g. Fee et al., 1996a; Harris et al., 2000). Our model assumes that most of the S.D. excess can be accounted for by cause (1).

3.1.2. Noise statistical properties

The noise auto- and cross-correlation functions (Fig. 1E) were obtained as described in the methods. Note their decay time is typically less than 1 ms. This decay time is similar to the event duration partly because the noise contains many spikes, too small to be detected as events.

3.1.3. Validation of the second-order noise description

It is clear that given a time dependent signal, one can always compute a correlation function (a single autocorrelation function if one records a single channel and several auto- and cross-correlation functions if one records several channels). We compute these functions in order to build the noise covariance matrix and we assume that this is a (relatively) complete description of the noise statistics. Nevertheless, one can imagine

Fig. 4. A, origin of the sampling jitter illustrated with simulated data. A1, two events are obtained by sampling an ideal waveform with two different origins. The ideal waveform is 450-points long and the two samples are 45-points long. The scale bars are arbitrary and are just labeled to help comparison with real data. A2, the two sampled events are aligned on their peaks and have slightly different time courses. A3, one hundred such events were generated from the ideal waveform (see text). Black line, mean event; grey line, S.D. The dashed line is the zero SD level (no noise was added to this simulation). B, cancellation of the sampling jitter illustrated on event 2. B1, event 2 (diamonds) is exactly interpolated with a sinc function whose period is twice the sampling period (black trace, the peak value is 1). Continuous grey line at zero. B2, interpolated event 2 needs to be slightly shifted to overlap exactly with the ideal waveform. The required shift is easily obtained by minimizing the χ^2 (see methods). C, D and E, sampling jitter cancellation on real data. C1, 139 events originating from unit 3 (data in Fig. 1) on site 1, before sampling jitter cancellation. Vertical bar, 100 μV. C2, mean event (top) and S.D. (bottom). Dotted line, S.D. level expected from the noise properties (15.4 μV, see Fig. 1C and E). Vertical bar, 10 μV (applies only to the SD trace on C2 and D2). The SD increase is slightly less pronounced than on the simulated case (A3) because some background noise is present and the total SD (SD_{Total}, which is the one displayed) is equal to ($\sqrt{\text{S.D.}^2_{\text{Noise}} + \text{S.D.}^2_{\text{Jitter}}}$). D1, the same 139 events after sampling jitter cancellation. D2, mean event (top) and S.D. (bottom). The horizontal scale bar applies to the four graphs, C1, C2, D1 and D2. E, χ^2 distributions before (thin line, $\langle\chi^2\rangle$) and after (thick line, $\langle\chi^2\rangle = 168$) sampling jitter cancellation (performed simultaneously on the four recording sites). Dotted line, noise χ^2 distribution ($\langle\chi^2\rangle = 171$).

Fig. 5. A, black curve, empirical Mahalanobis distance distribution obtained from a noise sample with 2000 events after noise whitening (see methods), the expected χ^2 distribution is shown in red. B, χ^2 PDFs. Red, expected PDF; black histogram, empirical probability density estimate from the same noise sample as in A and after coordinate transformation. C, third moment distribution of the whitened noise (see text).

plausible scenarios where this would not be the case. If, for example, the background noise is non-stationary (e.g. Fee et al., 1996a), several noise covariance matrices could be required successively to describe the noise, while a single one could not be an accurate model. Alternatively, the noise could be stationary, but with third- or higher-order moments.

One way to test that the covariance matrix is a full description of the noise is to see how well the Mahalanobis distance distribution fits the χ^2 distribution, as described in Section 2. Fig. 5A,B illustrate the empirical cumulative distribution (5A) and density (5B) of Mahalanobis distances for an actual noise sample. The expected values for these quantities have been plotted as well, the close match between actual and expected entities suggest that the noise distribution is well approximated by a multivariate Gaussian distribution.

Although the Mahalanobis distance test is quite sensitive, it is a necessary but not sufficient test of the accuracy of our noise description. We thus performed an additional test aimed at detecting deviations of the actual noise from its representation based on its covariance matrix. We estimated the distribution of the third moment about the mean of noise sample (see Section 2). Fig. 5C shows that for 500 randomly chosen coordinate triplets the average value of the third moment (sample size = 2000) has a Gaussian distribution with 0 mean and an S.D. of $1/\sqrt{2000}$, as expected. Taken together these results suggest that a noise description based on its covariance matrix is accurate enough for our purpose.

3.1.4. Application to real data

The methods described can now be applied to the real data of Fig. 1. Once the specific data generation model has been obtained (see Section 2), it is used to classify each of the 1493 events detected during 60 s of continuous data acquisition. Of the 1493 events detected, 1361 were classified as pure events (294 events from unit 1, 391 from unit 2, 139 from unit 3, 333 from unit 4 and 204 from unit 5), 89 were classified as superpositions of two different units and 43 (i.e. less than 3%) were classified as outliers.

The pure events of three of the five units are displayed in Fig. 6B1–3 together with their mean and S.D. The S.D. test seems to be met by the events of units 3 (Fig. 6B1) and 4 (Fig. 6B2) but not by the events of unit 5, which do not have a flat S.D. (Fig. 6B3). This is confirmed by the χ^2 test (Fig. 6C). The χ^2 distributions are expected to be on the left side on the noise χ^2 distribution (dotted line, Fig. 6C) for two reasons: they are computed by using the mean event of each unit as a template (i.e. 1 degree of freedom is lost) and the sampling jitter cancellation results in a slight overfit (see Fig. 4D2). On this basis, the distributions of units 2 and 5 are suspect.

The projection test (Fig. 6D) confirms the poorer quality of units 2 and 5 but shows that all units can be unambiguously distinguished. The empirical projection density estimates have been plotted on each graph (histograms) together with the expected distributions (colored curves). The projections are obtained by projecting all the pure events generated by one or the other unit of the pair. The expectations are completely defined by the knowledge of the distance between the two units of a pair and by the respective number of pure events generated by each of the two units (e.g. in the top graph, the distance between the two units is 8.55, the number of events from unit 1 is 294 and the number of events from unit 2 is 391; therefore, the integral of the blue Gaussian is 294/(294 + 391) = 0.43 and the integral of the green Gaussian is 0.57). Based on those tests, spikes originating from units 1, 3 and 4 would be kept for further analysis while spikes from units 2 and 5 would be discarded or at least taken with caution, for each of these two distributions likely contains more than one unit.

4. Discussion

We have shown that a very simple model can ‘explain’ electrophysiological data collected by extracellular recordings in the locust AL. The combination of an accurate noise model with an explicit model for data generation leads to specific quantitative tests that the classified data should meet. These tests are objective and can be graphically displayed, thus enabling the experimenter or the reader to assess the quality or trustworthiness of the analyzed data. It should be clear that the tests can be applied to the final results of any classification procedure.[6] These tests could therefore form a basis for comparison between different spike sorting techniques. The less rigorous ‘cluster cutting’ methods, used in particular by commercial software, sometimes leave the user or the reader with untestable confidence in the data. The adoption of objective measures such as those we propose here would, we believe, help alleviate this growing problem.

Our method does not take spike timing into account at any stage. That is, no explicit refractory period is set, and no general form of the spike train autocorrelation is required, as it is in many other methods (e.g. Fee et al., 1996b; Harris et al., 2000). In this way, spike timing information can be used as another, independent

[6] Two of the tests require a knowledge (or at least an estimate) of the noise covariance matrix Γ, but that can be estimated from the classified events.

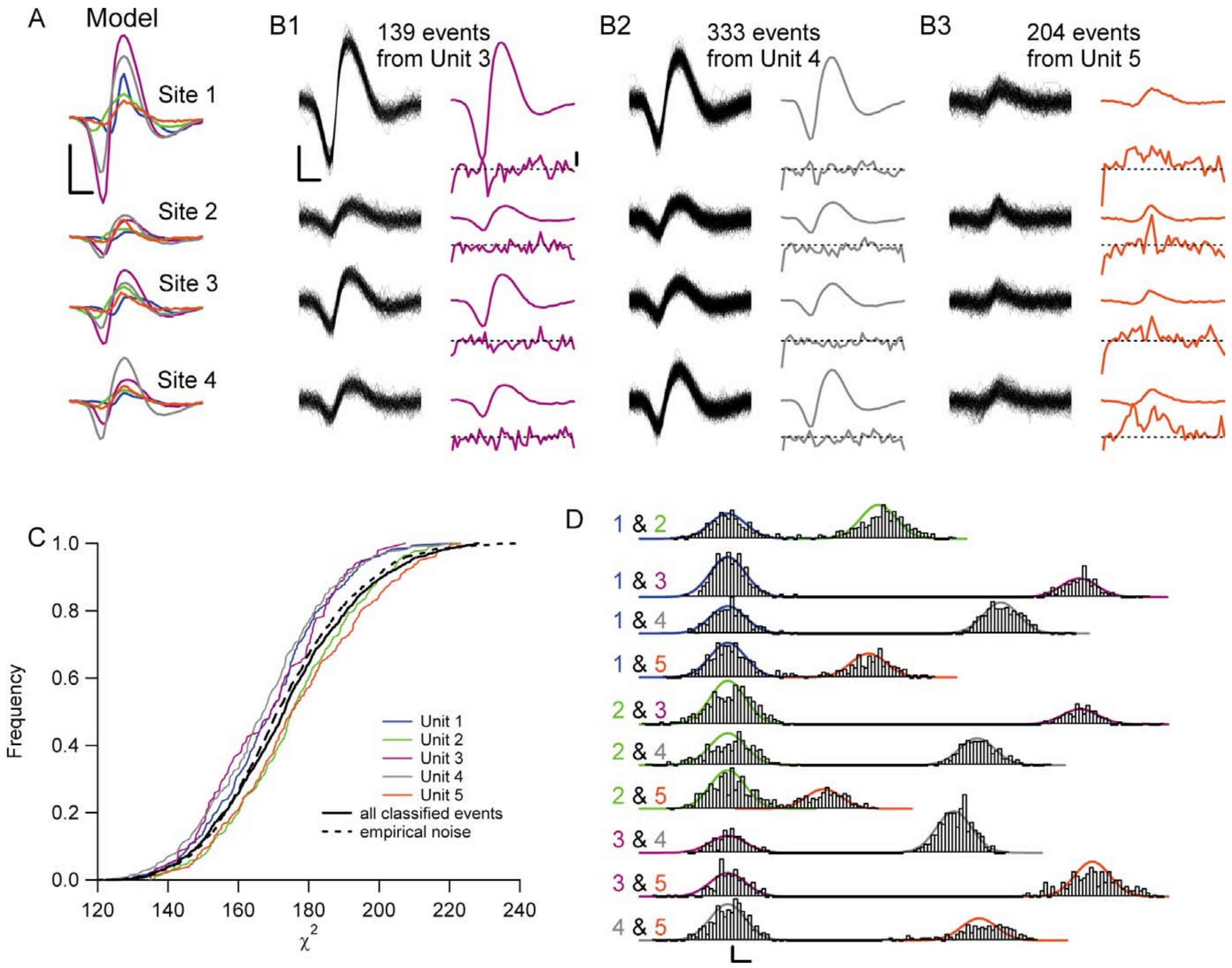

Fig. 6. Example of classification on real data from the locust antennal lobe. A, waveforms of each of the five units of the model on each of the four recording sites. Vertical, 100 μV; horizontal, 0.5 ms. B, events from three of the five units. B1 left, the 139 events generated by unit 3; B1 right, mean event and S.D. Calibrations: left, as in A; right, S.D. of site 1, 20 μV. Dotted lines on S.D. graphs: expected noise (see Fig. 1C). B2 and B3, same as B1, for units 4 and 5. All graphs in B have same scale. C, χ^2 test. Cumulative distributions of the pure events from the five units and from all the 1450 classified events (pure events from each of the five units and 89 events classified as superpositions). Dashed line, empirical noise χ^2 distribution. D, Projection test. For each of the ten possible pairs of units, the empirical projection of the events belonging to one or the other unit of the pair is shown (histograms). The expected probability densities are shown superposed on each graph (see text). Vertical bar, 0.1. Horizontal bar, 1. Note the absence of overlap between the projections of any two units, as well as the systematic discrepancy between the empirical and expected densities for units 2 and 5.

measure of quality. In our data we find very reproducible inter-spike interval distributions, among neurons in a given animal, as well as across animals (C. Pouzat, O. Mazor and G. Laurent, in preparation), further confirming the accuracy of the procedure. Furthermore, our tests routinely detect situations which would be missed by timing-based test alone. For example, consider a cluster that contains only half of the spikes generated by a single unit, either because that unit was split between two clusters or because a substantial percentage of its spikes were not detected. While this cluster will still exhibit a 'normal'-looking refractory period and autocorrelation, it should still be detected as incomplete by the tests we propose (by either the projection test or the S.D. test, respectively).

The model we use here for the locust data is clearly the simplest possible one. It is not expected to hold for all data sets; however, we expect it to work successfully in a wide variety of experimental conditions. In rat neocortex, for example, Fee et al. find that for most of the units they recorded "[t]he variability of spike residuals is nearly identical with that of background activity" (1996a, p. 3831, see also Fig. 1b,d,e and 2 in the same article), indicating that the tests we propose should work in this system as well. Furthermore, using the framework we describe, one can introduce more sophisticated models of data generation to analyze more complicated data sets. Once a new model is specified, the statistical tests we introduced can be readily generalized. Of particular interest are spikes with non stationary waveforms (e.g. within bursts). One way to model the waveforms of spikes in such neurons would be to assign to each unit a cluster-specific covariance matrix, to account for the added variability in spike shape. These cluster specific matrices would complement a global covariance matrix describing the noise. Although the use of a covariance matrix to model spike waveform variability is only an approximation, preliminary results using in vivo data collected from the rat hippocampus (generously shared by K. Harris and G. Buzsaki) indicate that the distribution of spike waveforms from a bursty cell is well described by its second order statistics (i.e. by a covariance matrix). Another alternative would be to develop a model of the dependence of the spike waveform on the inter-spike interval (as suggested by Fee et al., 1996a) and use this model to scale the template before computing the residual. Our tests would then be directly applicable.

Vertebrate data also often exhibit non stationary noise (e.g. Fee et al., 1996a). Such data would require a more precise description of the noise. For instance, an extension of the current model could use a time dependent noise covariance matrix. The noise whitening would then be applied by taking into account each event's time of occurrence.

Acknowledgements

We thank Dr Michael Roukes for his help with the silicon probes, Drs Kenneth Harris and Gyorgy Buzsaki for sharing their hippocampus data, Dr Fabrizio Gabbiani for discussion, Drs Mark Stopfer, Jean Diebolt and Pierre Brémaud for comments, Dr Gilles Celeux for suggesting the use of the Bayesian Information Criterion and the developers teams of Scilab and Mixmod. Supported by NIDCD, the Keck and McKnight Foundations (GL), and a Department of Defense NDSEG Fellowship (OM). Multichannel silicon probes were kindly provided by the University of Michigan Center for Neural Communication Technology, sponsored by NIH NCRR grant P41RR09754.

References

Atiya AF. Recognition of multi-unit neural signals. IEEE Trans Biomed Eng 1992;39:723–9.

Biernacki C, Govaert G. Choosing models in model-based clustering and discriminant analysis. INRIA Research Report #3509. http://www.inria.fr/rrrt/rr-3509.html, 1998. 1–14.

Bishop CM. Neural Networks for Pattern Recognition. Oxford: Clarendon, 1995.

Boyles RA. On the convergence of the EM algorithm. J R Stat Soc B 1983;45:47–50.

Brandt S. Data Analysis. New York: Springer-Verlag, 1999.

Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B 1977;39:1–38.

Drake KL, Wise KD, Farraye J, Anderson DJ, Bement SL. Performance of planar multisite microprobes in recording extracellular single-unit intracortical activity. IEEE Trans Biomed Eng 1988;35:719–32.

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

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 1996b;69:175–88.

Fraley C, Raftery AE. How many clusters? Which clustering method? Answers via model-based cluster analysis. Comput J 1998;41:578–88.

Frostig RD, Lieke EE, Tso DY, Grinvald A. Cortical functional architecture and local coupling between neuronal activity and the microcirculation revealed by in vivo high resolution optical imaging of intrinsic signals. Proc Natl Acad Sci 1990;87:6082–6.

Glaser EM, Marks WB. Online separation of interleaved neuronal pulse sequences. Data Acquisition Process Biol Med 1968;5:137–56.

Harris KD, Henze DA, Csicsvari J, Hirase H, Buzsaki G. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. J Neurophysiol 2000;84:401–14.

Hulata E, Segev R, Shapira Y, Benveniste M, Ben-Jacob E. Detection and sorting of neural spikes using wavelet packets. Phys Rev Lett 2000;85:4637–40.

Jack JJ, Redman SJ, Wong K. The components of synaptic potentials evoked in cat spinal motoneurones by impulses in single group Ia afferents. J Physiol (Lond) 1981;321:65–96.

Laurent G, Davidowitz H. Encoding of olfactory information with oscillating neural assemblies. Science 1994;265:1872–5.

Laurent G, Naraghi M. Odorant-induced oscillations in the mushroom bodies of the locust. J Neurosci 1994;14:2993–3004.

Letelier JC, Weber PP. Spike sorting based on discrete wavelet transform coefficients. J Neurosci Methods 2000;101:93–106.

Lewicki MS. Bayesian modeling and classification of neural signals. Neural Comput 1994;6:1005–30.

Lewicki MS. A review of methods for spike sorting: the detection and classification of neuronal action potentials. Network 1998;9:R53–78.

Ling L, Tolhurst DJ. Recovering the parameters of finite mixtures of normal distributions from noisy record: an empirical comparison of different estimating procedures. J Neurosci Methods 1983;8:309–33.

McLachlan GJ, Krishnan T. The EM Algorithm and Extensions. New York: Wiley, 1997.

Millecchia R, McIntyre R. Automatic nerve impulse identification and separation. Comput Biomed Res 1978;11:459–68.

Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proc Natl Acad Sci 1992;89:5951–5.

Papoulis A. Circuits and Systems. A Modern Approach. New York: Holt, Rinehart and Winston, 1980.

Redner RA, Walker HF. Mixture densities, maximum likelihood and the EM algorithm. SIAM Rev 1984;26:195–239.

Sahani M. Latent variable models for neural data analysis. PhD Thesis, California Institute of Technology, Pasadena, 1999.

Schwarz G. Estimating the dimension of a model. Ann Stat 1978;6:461–4.

Wehr M, Pezaris JS, Sahani M. Simultaneous paired intracellular and tetrode recordings for evaluating the performance of spike sorting algorithms. Neurocomputing 1999;26–27:1061–8.

Wu CFJ. On the convergence properties of the EM algorithm. Ann Stat 1983;11:95–103.

Wu JY, Cohen LB, Falk CX. Neuronal activity during different behaviors in Aplysia: a distributed organization? Science 1994;263:820–3.

Zhu Z, Lin K, Kasamatsu T. Artifactual synchrony via capacitance coupling in multi-electrode recording from cat striate cortex. J Neurosci Methods 2002;115:45–53.