

Robust Unsupervised Detection of Action Potentials With Probabilistic Models

Raul Benitez*, *Member, IEEE*, and Zoran Nenadic, *Member, IEEE*

Abstract—We develop a robust and fully unsupervised algorithm for the detection of action potentials from extracellularly recorded data. Using the continuous wavelet transform allied to probabilistic mixture models and Bayesian probability theory, the detection of action potentials is posed as a model selection problem. Our technique provides a robust performance over a wide range of simulated conditions, and compares favorably to selected supervised and unsupervised detection techniques.

Index Terms—Action potentials, Bayesian probability theory, continuous wavelet transform, expectation maximization algorithm, finite mixture models, maximum likelihood principle, receiver operating characteristic, unsupervised detection.

I. INTRODUCTION

IN response to sensory stimuli and/or cognitive processes, the majority of cortical neurons generate a sequence of action potentials (APs), commonly known as spikes. While the nature of neural code is not yet fully understood, there is convincing evidence that the temporal features of spikes, e.g., their firing rates [1], [2], or arrival times [3], are involved in the encoding, transmission, and processing of information within the nervous system. In many behavioral studies, and *in vivo* experiments in general, APs are measured extracellularly by placing a high-impedance microelectrode in the vicinity of electrically active neurons [4]–[9]. If the features of spike trains can be linked to sensory, motor or cognitive variables, a great deal of information can be learned about the brain function. On the other hand, the signals measured by the electrode are inevitably corrupted by noise, generated by many sources, most notably by the microelectrode itself (thermal noise) [6], [7], [10], the activity of distant neurons (biological noise) [10], [11], and the intrinsic variability of neurons' ionic channels (channel noise) [12]. Therefore, the first step in the processing of any extracellularly recorded data is the detection of individual spikes and their separation from the noisy background. Any errors during this preliminary data processing step will necessarily propagate through all subsequent analyses. Therefore, to ensure correct scientific conclusions, it is ultimately important to accurately detect and localize the occurrence of APs within the extracellularly recorded signals.

Manuscript received February 17, 2007; revised September 22, 2007. *Asterisk indicates corresponding author.*

*R. Benitez was with the Department of Biomedical Engineering, University of California, Irvine, CA 92697 USA. He is now with the Department of Automatic Control, Universitat Politècnica de Catalunya, Compte Urgell 187, 08036 Barcelona, Spain (e-mail: raul.benitez@upc.edu).

Z. Nenadic is with the Department of Biomedical Engineering and the Department of Electrical Engineering and Computer Science, University of California, Irvine, CA 92697 USA (e-mail: znenadic@uci.edu).

Digital Object Identifier 10.1109/TBME.2007.912433

Due to its importance to experimental neuroscience, detection of APs in noisy background is a well-researched area. Early contributions include the use of amplitude thresholding [13], principal component analysis [14], discrete wavelet transform (DWT) [15], neural networks [16], power detection [17], template matching [18], etc. While all of these techniques can be automated and therefore are amenable to computer implementations, a vast majority of these methods rely on some level of human supervision and intervention, such as, for example, a separation of signal and noise based on some preliminary data. This *training set* is then used to estimate the parameters of the detection statistic, calculate the principal components, train a neural network, construct spike templates, etc. Truly unsupervised, automated spike detection techniques, however, have not received as much attention, primarily because extracellularly recorded neural data is traditionally analyzed off-line and under the close supervision of the human operator.

The development of a robust, *fully unsupervised* spike detection algorithm may be beneficial to many research areas (e.g., experimental neuroscience, neurophysiology, neural engineering) for several reasons. First, with the recent advances in microelectrode array technology that have enabled simultaneous recording from a large number of neurons [19]–[21], the traditional paradigm of manual tuning of detection parameters becomes tedious and time consuming. As this process is often iterative, it may also be highly impractical, especially for arrays with a large number of recording channels. Second, in the absence of the “ground truth,” the human operators must rely on their intuition and experience to set the detection parameters, which could impose a sampling bias, especially under low signal-to-noise ratio (SNR) conditions. Third, robust and fully unsupervised signal processing tools, including spike detection, will play an important role in the future brain-machine interfaces (BMIs), which will require more autonomous operation and gradual human operator disengagement. Some examples of autonomous functions in the context of acute extracellular recordings have been recently demonstrated in [22] and [23]. Finally, due to inherent nonstationarities of neurophysiologic processes and signals, the detector parameters often require periodic readjustments, which may be time consuming if human involvement is needed. Robust, unsupervised spike detection algorithms would eliminate the role of the human operator, while providing consistent performance under nonstationary recording conditions.

Based on the continuous wavelet transform (CWT) and ideas from Bayesian probability theory, we pose the problem of AP detection within a probabilistic modeling framework. We then proceed to estimate the parameters of such a model using the maximum likelihood (ML) principle, and we show how this information can be used to separate the signal and noise in a fully

unsupervised fashion. Probabilistic models have been used for unsupervised spike classification [24], [25]. However, before APs are classified, they need to be detected, and to the best of our knowledge, the use of probabilistic models for unsupervised AP detection is a novel idea. The details of our method are given in Section II. In Section III we validate the performance on our method on a set of data synthesized using signals recorded by an array of microelectrodes. We also compare the performance of our method against those of other unsupervised techniques. The performances and some peculiarities of AP detection are discussed in Section IV. Final conclusions are given in Section V, and some mathematical derivations are given in the Appendix.

II. SPIKE DETECTION WITH PROBABILISTIC MODELS

Our detection algorithm consists of three steps, namely the following:

- 1) extract features from data via the continuous wavelet transform at selected scales;
- 2) estimate the parameters of the probabilistic model of the features;
- 3) estimate the arrival times of individual APs.

Each of the steps will be thoroughly explained in the subsequent sections. We also note that steps 2) and 3) are compatible with various feature extraction techniques, and that the use of CWT for feature extraction is not mandatory. However, we argue below that CWT is well-suited for unsupervised AP detection.

A. Feature Extraction With the Continuous Wavelet Transform

Classical detection theory shows that under the additive noise model, the SNR-optimal detection is achieved by a *matched filter* [26]. The synthesis of the matched filter, however, requires the knowledge of the signal to be detected. Therefore, in unsupervised applications, the SNR-optimal detection is not feasible. An alternative strategy is to approximate the matched filter by a template, obtained by averaging multiple APs from a single well-isolated neuron [18]. Since the synthesis of the template requires a substantial human supervision, this technique is not amenable to unsupervised applications. When the signal to be detected is not known, which is the case in unsupervised AP detection, the detector performance may be representation dependent [27]. Wavelet functions of compact support are well-suited to representing signals that contain transient phenomena [28], [29], such as APs.

1) *Choice of Wavelet Family*: It is important to choose, among a wide variety of existing wavelet families, those that are suitable to AP detection. Our choice is motivated by the shapes of typical APs, and a more detailed discussion on this topic can be found in [30]. Essentially, there exist wavelet families from biorthogonal class [28], whose basic shapes are reminiscent of the bi- or tri-phasic shape of a typical AP. (The biophysical determinants of the shape of extracellular APs are discussed in [7] and [31].) These wavelets, such as *bior 1.3*¹ and *bior 1.5*, are defined via splines and do not have an explicit analytic expression [28]. Nevertheless, they are amenable to fast filtering algorithms [32] and can be efficiently implemented in software. By convolving the neural signal with such

¹The wavelet notation throughout this article is consistent with that of MATLAB Wavelet Toolbox.

a wavelet filter, the coefficients corresponding to noise appear as zero-mean random fluctuations, while the coefficients of APs deviate significantly from zero, and are considered outliers. Due to the spike-like shape of these wavelets, the AP energy is distributed over few coefficients (sparse representation), while the noise energy is uniformly distributed over all coefficients. It is then easier to separate the AP and noise in the wavelet domain, and this is the main idea behind our approach, the details of which will be presented in Section II-B.

2) *Choice of Wavelet Scale*: Similar to choosing the wavelet family, biophysical considerations can be used to choose the wavelet scales. Wavelet coefficients at fine scales are mostly due to signal discontinuities and high frequency noise [29], whereas the coefficients at coarse scales represent low-frequency signal variations (e.g., local field potentials), and are irrelevant for AP detection [30]. Based on the common range of duration of APs (typically from 0.5 to 1.5 ms [7], [8]), the set of wavelet scales can be significantly restricted. These scales, denoted by $\mathcal{A} = \{a_1, a_2, \dots, a_J\}$, will be referred to as *relevant scales* throughout this paper. Scale restriction represents an important difference between our approach and other wavelet-based approaches to AP detection [15], [33] and signal detection [34], [35], in general, which are based on DWT and dyadic scaling. Restricting the set of scales also reduces the computational cost of our method.

3) *Translation Invariance*: The detectability of an AP should not depend on its position within the analyzed time series [36]. The commonly used DWT does not have the desirable translation invariance property [29], and modifications of the DWT algorithm exist that give rise to the stationary wavelet transform (SWT) (e.g., [37]). However, SWT also operates on dyadic scales and the translation invariance of SWT is only approximate. For these reasons, we propose the use of CWT as a suitable tool for AP detection. The suitability of CWT for AP detection has been discussed at length in [30]. Fig. 1 shows a two-dimensional (2-D) feature representation of extracellularly recorded data,² obtained by applying *bior 1.3* CWT at scales $\mathcal{A} = \{a_1, a_2\}$, corresponding to 0.5 and 1.5 ms, respectively. Each feature vector, $\mathbf{x} \in \mathbb{R}^2$, consists of wavelet coefficients $X(a_1, b)$ and $X(a_2, b)$, where b is the translation of the wavelet. The coefficient are calculated by projecting the signal x onto the wavelet basis $\psi_{a,b}$

$$X(a, b) \triangleq \int_{\mathbb{R}} x(t) \psi_{a,b}(t) dt \quad (1)$$

where the basis functions are obtained from the mother wavelet, ψ , by scaling and translation, i.e.,

$$\psi_{a,b}(t) \triangleq \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), \quad a, b \in \mathbb{R}$$

For a discrete signal $x = \{x[0], x[1], \dots, x[N-1]\}$, the set of translations is discrete, i.e., $b \in \mathcal{B} \triangleq \{0, 1, \dots, N-1\}$, in

²An array of 64 glass-coated platinum-iridium microelectrodes (Microprobe Inc. Gaithersburg, MD) with the impedance of 300 k Ω was chronically implanted in the posterior parietal cortex of a Rhesus monkey (*Macaca Mulatta*). Broadband signals (1 Hz-10 kHz) were acquired using a multichannel data acquisition system (Plexon Inc., Dallas, TX), and digitized (sampling rate: 20-KHz, resolution: 12-bit) with a data acquisition card (National Instruments, Austin, TX).

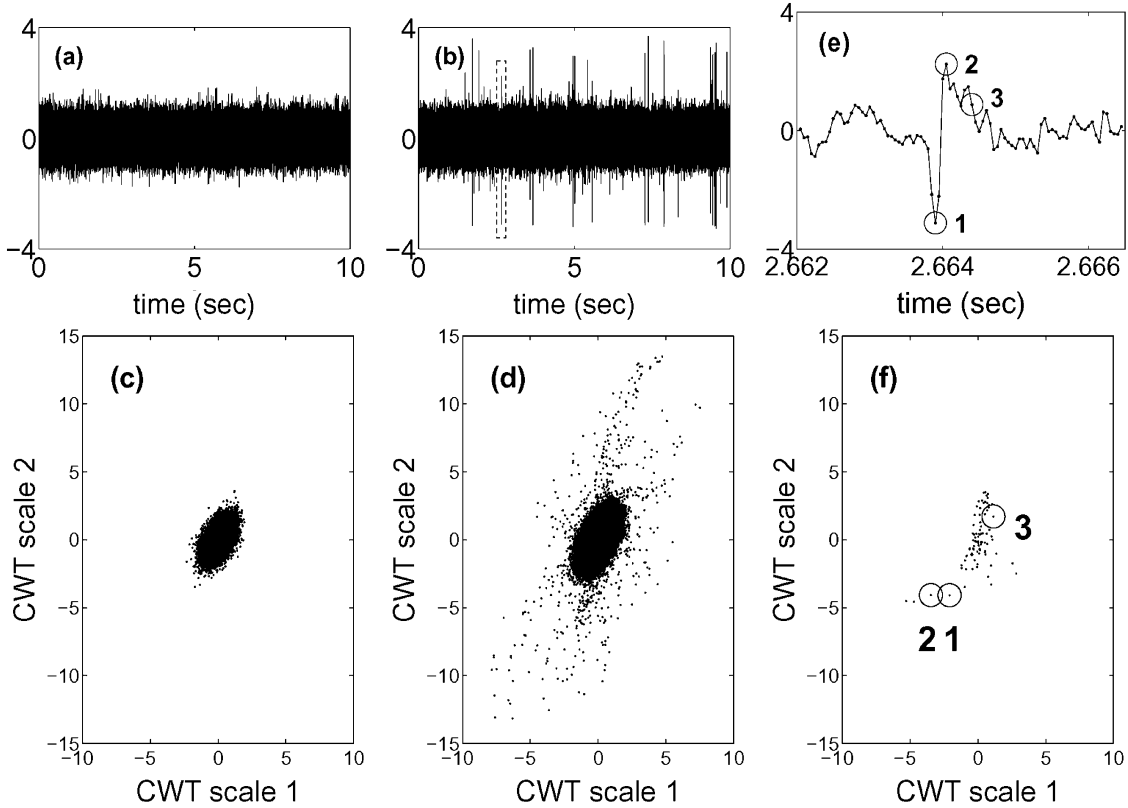


Fig. 1. (a) Recorded data with no distinct APs—noise (see Section III-A). (b) Data with prominent APs. (c) Continuous wavelet representation of data shown in (a), with $\mathcal{A} = \{a_1, a_2\}$. The wavelet function is from *bior 1.3* family of wavelets [28]. $\text{BIC}_1 = -6.0086 \times 10^5$, $\text{BIC}_2 = -6.0089 \times 10^5$ (see Section II-C for explanation). (d) Equivalent plot for data shown in (b). $\text{BIC}_1 = -6.5126 \times 10^5$ and $\text{BIC}_2 = -6.3314 \times 10^5$. (e) An AP from the signal in (b) with three samples marked by 1, 2, and 3. (f) Representation of signal in (e) in the CWT feature space. Note that each sample of the signal is represented by a point in the feature space.

which case (1) is replaced by $X(a, b) \triangleq \langle x, \psi_{a,b} \rangle$, where $\langle \cdot, \cdot \rangle$ stands for the inner product in \mathbb{R}^N . It follows readily that if the observation sequence, x , contains only noise, w , i.e.,

$$x[t] = w[t], \quad t \in \mathcal{B}$$

then the wavelet coefficients $\{X(a, b) : a \in \mathcal{A}, b \in \mathcal{B}\}$ will represent the samples of zero-mean random variables [30] [see Fig. 1(c)]. If a spike s is present during some transient period \mathcal{B}_0 , i.e.,

$$x[t] = s[t] + w[t], \quad t \in \mathcal{B}_0$$

the samples $\{X(a, b) : a \in \mathcal{A}, b \in \mathcal{B}_0\}$ will deviate from zero. The deviations will be largest if the set of relevant scales \mathcal{A} is chosen such that the support of $\{\psi_{a,b} : a \in \mathcal{A}\}$ matches the duration of APs. These statistical considerations, therefore, suggest that the cluster centered at the origin is likely representing the noise, while the feature points outside of the central core are likely due to spikes [see Fig. 1(d)]. Note that unlike its discrete counterpart, CWT yields a redundant representation (more basis functions than required), and that each time sample is represented by a point in the feature space [Fig. 1(e), (f)]. Also note that CWT is capable of dealing with APs with asymmetrical peaks, as well as with APs of reversed polarities. An illustrative example is shown in Fig. 1, where both the positive and negative peaks of AP in (e) qualify as signals in the feature space (f). Finally, the symmetry properties of the CWT ensure that two

signals with reversed polarities are represented as specular images in the feature space.

B. Estimation of Probabilistic Models

1) *Finite Mixture Models*: When a sampling distribution of statistical data is complex, it is often modeled using the finite mixture approach [38]. In particular, Gaussian mixture models [39] have been used in many applications involving density estimation and related cluster analysis [40], [41]. More specifically, applications of Gaussian mixtures to modeling of APs can be found in [23], [25]. In the context of signal detection, Gaussian mixture models take a very simple form. If a feature $\mathbf{x}_i \in \mathbb{R}^J$ is generated by the noise, we propose a single-component probability density function (PDF) model

$$\mathcal{M}_1 : f(\mathbf{x}_i | \theta_1) = \mathcal{G}(\mathbf{x}_i, \theta_1)$$

where \mathcal{G} is a Gaussian PDF, $\theta_1 \triangleq \{\mu_1, \Sigma_1\}$, and μ_1 and Σ_1 are the mean and covariance matrix, respectively. If \mathbf{x}_i contains both the signal and noise, a convex combination of two components is proposed, i.e.,

$$\mathcal{M}_2 : f(\mathbf{x}_i | V, \tau, \theta_2) = \tau_1 \mathcal{U}(\mathbf{x}_i, V) + \tau_2 \mathcal{G}(\mathbf{x}_i, \theta_2)$$

where $\mathcal{U}(\mathbf{x}_i, V) = (1)/V$ is a uniform PDF, V is the hypervolume of the data, $\tau \triangleq \{\tau_1, \tau_2\}$, with $\tau_1 + \tau_2 = 1$, and $\theta_2 \triangleq \{\mu_2, \Sigma_2\}$. In the Gaussian mixture framework, a uniform component is often used as a model for outliers [40], [41]. Thus,

from the noise perspective, signals are treated as *outliers*, and the signal detection problem reduces to the question whether the outliers are present in the noise.

C. Model Selection and Parameter Estimation

We proceed by answering the following questions, given a sequence of features (samples) $\mathcal{X} = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{N-1}\}$:

- 1) Which model, \mathcal{M}_1 or \mathcal{M}_2 , fits the data \mathcal{X} better?
- 2) What are the parameters of the optimal model \mathcal{M}^* ?

The first question is answered through a model selection process, which establishes the absence/presence of APs. The second problem can be solved via the ML principle. We recast the first question using the Bayesian probability framework [42], i.e., we calculate the (posterior) probability, $p(\mathcal{M}_k|\mathcal{X}, \mathcal{I})$ of each model given the features, \mathcal{X} , and some prior information, \mathcal{I} .³ We then propose to choose the model with a higher posterior probability, which after invoking Bayes' theorem

$$p(\mathcal{M}_k|\mathcal{X}, \mathcal{I}) = \frac{f(\mathcal{X}|\mathcal{M}_k, \mathcal{I})p(\mathcal{M}_k|\mathcal{I})}{f(\mathcal{X}|\mathcal{I})}, \quad k = 1, 2$$

and assuming equal prior probabilities $p(\mathcal{M}_k|\mathcal{I}) = (1)/(2)$ yields

$$\text{BF} \triangleq \frac{f(\mathcal{X}|\mathcal{M}_1, \mathcal{I})}{f(\mathcal{X}|\mathcal{M}_2, \mathcal{I})} > 1.$$

Thus, the decision (choose \mathcal{M}_1 if $\text{BF} > 1$, and vice versa) is based on the ratio of integrated likelihoods, also known as the Bayes factor (BF) [43]. Unfortunately, even if models are simple, the integrated likelihood is not. For example

$$f(\mathcal{X}|\mathcal{M}_1, \mathcal{I}) = \prod_{i=0}^{N-1} \int \mathcal{G}(\mathbf{x}_i, \theta_1) \phi(\theta_1) d\theta_1$$

where $\phi(\theta_1)$ is a PDF of θ_1 . In general, the evaluation of this integral requires multi-dimensional numerical integration, and will not be pursued here. Instead, we use an approximation based on Bayesian information criterion (BIC) [44]. For other approaches to model selection problem, the reader is referred to [45]–[47]. BIC is defined as

$$\begin{aligned} \text{BIC}_k &\triangleq \log \mathcal{L}_k^*(\mathcal{X}) - \frac{\nu_k}{2} \log N \\ &\approx \log f(\mathcal{X}|\mathcal{M}_k, \mathcal{I}), \quad k = 1, 2 \end{aligned}$$

where \mathcal{L}_k^* and ν_k are the maximum likelihood and the number of independent parameters of model \mathcal{M}_k , respectively. A straightforward application of the ML principle to model \mathcal{M}_1 yields

$$\mathcal{L}_1^*(\mathcal{X}) = \prod_{i=0}^{N-1} \mathcal{G}(\mathbf{x}_i, \theta_1^*)$$

where the optimal parameters, θ_1^* , are the sample mean μ_1^* and the sample covariance Σ_1^* . To evaluate \mathcal{L}_2^* , more elaborate (numerical) ML schemes are needed, and we apply the widely

³Prior information \mathcal{I} represents what was known about the problem (e.g., spikes are triphasic and localized in time) before observing \mathcal{X} . In Bayesian probability theory, all probabilities must be conditioned upon \mathcal{I} .

used expectation-maximization (EM) algorithm [48] (see the Appendix).

Once \mathcal{L}_1^* and \mathcal{L}_2^* are known, the models can be compared based on their BICs. Choosing \mathcal{M}_1 as the optimal model indicates that no spikes are present in data. An example of this situation is illustrated in Fig. 1(c), where $\text{BIC}_1 > \text{BIC}_2$. If, on the other hand, \mathcal{M}_2 turns out to be optimal, APs are present [e.g., Fig. 1(d)], and they need to be separated from the noise. This can be achieved using the maximum posterior probability principle: assign \mathbf{x}_i to the signal, if $P(\mathbf{x}_i \in \mathcal{U}|\mathbf{x}_i, V^*, \tau^*, \theta_2^*) > P(\mathbf{x}_i \in \mathcal{G}|\mathbf{x}_i, V^*, \tau^*, \theta_2^*)$, and vice versa. Practically, this reduces to (see (6) in the Appendix)

$$\mathbf{x}_i \in \begin{cases} \text{signal,} & \text{if } \tau_1^* \mathcal{U}(\mathbf{x}_i, V^*) > \tau_2^* \mathcal{G}(\mathbf{x}_i, \theta_2^*) \\ \text{noise,} & \text{otherwise} \end{cases}.$$

D. Arrival Time Estimation

The separation of signal and noise in the feature space is necessary, but not sufficient to localize⁴ individual APs. For example, an AP sampled at 20 kHz and lasting 1 ms will be represented by 20 points in the feature space [see Fig. 1(e) and (f)]. When the AP is detected, some of its feature points will be classified as outliers (signal), while others (especially features representing the tails of APs) will be classified as noise. Ideally, the features of APs classified as signal should be contiguous in time, in which case it is easy to localize the spikes, e.g., by taking the time location of the largest⁵ feature vector, or by averaging the time location of the first and the last feature vector within this contiguous segment. The latter strategy is more robust with respect to noise, and was applied in this study. In many cases, however, the signal features of an APs will be split over two (or more) temporally contiguous segments, which if not handled properly, will result in multiple detections of a single AP. However, these disjoint contiguous segments are typically very close in time, and may be safely merged into a single segment. The arrival times are then estimated based on the merged segments. This type of postprocessing is performed sequentially from the onset of the signal for two contiguous segments at a time. The intermediate results are updated and the process is repeated until no further merging is possible. The tradeoffs of this process and recommendations regarding the choice of the merging threshold are discussed in [30].

III. EXPERIMENTAL RESULTS

To evaluate the performance of our detector analytically, a complete statistical description of the features under the models \mathcal{M}_1 and \mathcal{M}_2 is necessary. In practical applications, these statistics are never known, therefore the performance must be evaluated experimentally. On the other hand, experimentally recorded data is inadequate for validation purposes, primarily due to the absence of the ‘‘ground truth,’’ i.e., the inability to independently verify the number of putative APs and their arrival times. Furthermore, in experimental data it is hard to systematically control critical parameters of detector

⁴By localization we mean the association of a single AP to a single occurrence (arrival) time.

⁵A suitable norm for this space is the Mahalanobis norm (see the Appendix).

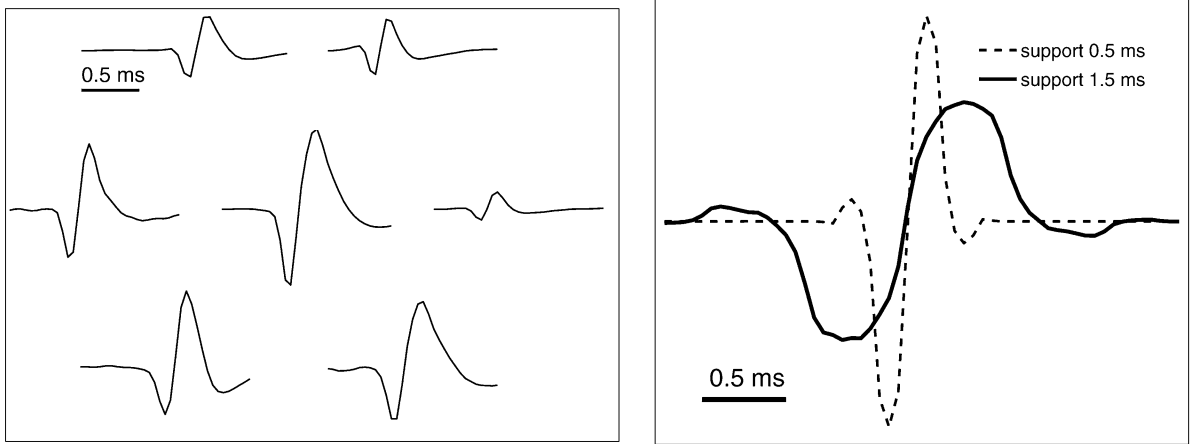


Fig. 2. Left panel: Seven spike templates identified from 400 s of recordings obtained from four channels of the multielectrode array. Principal component features were used during the classification procedure, and the quality of the templates was assessed by inspection of the residuals. Right panel: Waveform of the *bior 1.3* wavelets at scales corresponding to 0.5 and 1.5 ms. Note the similarity between the shapes of the wavelets and APs.

performance such as SNRs and firing rates. Finally, to ensure statistically significant results, a large number of trials (typically a few hundreds [26]) need to be performed, which is not feasible to achieve experimentally. Therefore, the validation of our detection algorithm will be performed on simulated data.

A. Modeling Neural Data

A modeling compromise between synthetic and experimental data is achieved by estimating templates from experimentally recorded APs, followed by a random temporal arrangement of the templates and addition of a suitably modeled noise. A similar modeling approach has been adopted in [30], [49], and [50], and many other studies concerning AP detection and classification.

1) *Construction of Spike Templates*: The templates were synthesized based on recordings from the monkey cortex (see Section II-A for recording details). A total of 400 s of high-SNR recordings from four channels was used. A short segment of one such recording is shown in Fig. 1(b). APs were detected in these data segments using a conservative amplitude threshold, set manually by the operator. Together with high SNR this reduced the probability of false alarm significantly, thereby ensuring that only genuine APs were collected in the process. The APs were then aligned using the maximum correlation method [30], and scatter plots were created using principal component features. Seven different AP clusters were visually identified and the spikes belonging to the same cluster were averaged to produce the templates. The quality of clustering was assessed by examining the residuals (defined as the difference between AP and the template) of APs within each cluster [11]. When necessary, APs with excessive residuals were upsampled and realigned, and the templates were re-estimated. While technically involved, the implementation of these steps is straightforward and the details can be found in [11]. Fig. 2 shows seven templates S_i obtained in the process. The template construction procedure was repeated with the continuous wavelet detection (CWD) [30], by imposing a high penalty of false alarm.

CWD yielded templates which were not distinguishable from the ones in Fig. 2, suggesting that the construction of templates did not impose any bias for further analysis.

2) *Construction of Spike Trains*: The arrival times of templates were generated by a homogeneous Poisson process with a refractory period of 2 ms enforced. The duration of each trial was set to 10 s, and the rate of the Poisson process was chosen to mimic low (10 Hz), medium (20 Hz), and high (40 Hz) firing rate (FR). To ensure multiunit spike trains are generated and avoid sampling bias, the template labels were drawn randomly from a uniform distribution, and the templates were normalized $\|S_i\|_\infty = 1$ ($i = 1, 2, \dots, 7$) and successively centered at the arrival times, thus forming *spike trains*. Note that the normalization of templates admits a unique definition of SNR for the whole spike train, i.e.,

$$\text{SNR} := \frac{\|S_i\|_\infty}{\sigma} = \frac{1}{\sigma} \quad (2)$$

where σ is the standard deviation of the noise. Neural noise was then taken from recordings that did not yield any identifiable spikes. A short segment of one such recording is shown in Fig. 1(a). Therefore, for each trial, a 10-s-long segment of neural noise was chosen at random from a data base. The noise was then scaled [according to (2) and SNR desired for each trial (3.5, 4.0, and 4.5 in this study)], and added to the spike train. The abundant volume of noise data allowed the construction of about 300 trials for each SNR-FR combination, without any noise overlap. The implications and advantages of this noise model are discussed at length in [23] and [30].

B. Performance Evaluation

For each SNR-FR combination, 300 Monte Carlo trials (10 s/trial) were constructed. In each trial the CWT features were extracted with *bior 1.3* wavelets whose scales corresponded to the support of 0.5 and 1.5 ms (see Fig. 2). The features were then modeled with \mathcal{M}_1 and \mathcal{M}_2 , estimated according to the procedure in Section II-C. The mixture \mathcal{M}_2 was consistently (for all SNR-FR combinations and all trials) the model with the highest BIC. This is not surprising, given that all the trials contained both the signal and noise. The points in the feature space belonging to the uniform component \mathcal{U} of model \mathcal{M}_2 were then considered parts of APs, and their arrival times were determined by means of the merging procedure described in

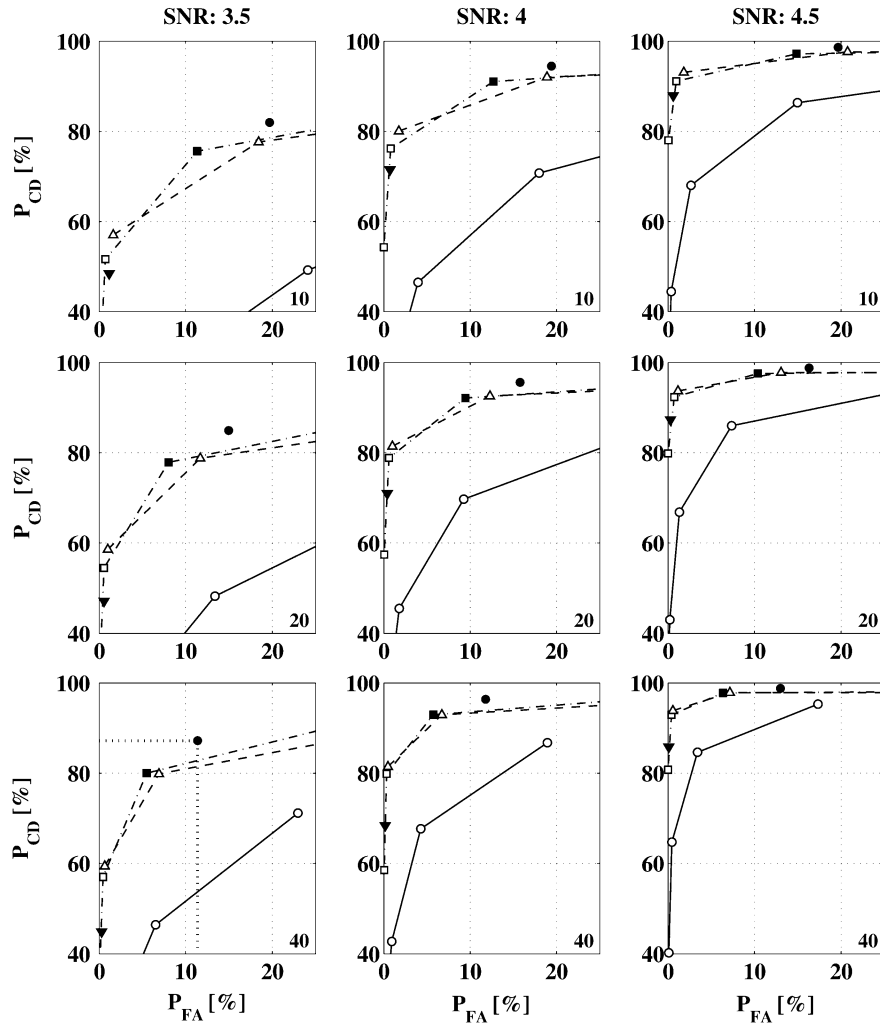


Fig. 3. ROC curves of various detection methods: PMD (\bullet), CWD (\square), UGT (\blacktriangledown), FDR (\triangle) and ATD (\circ). Each panel illustrates the performance for one SNR-FR combination (SNRs are given on top, and FRs are given in the right bottom corner).

Section II-D. The performance of our method, referred to as the probabilistic model detection (PMD) for brevity, was evaluated based on the average number of correct detections and false alarms over trials. A detection was declared correct if its estimated arrival time was within 0.5 ms from the true arrival time. Otherwise a false alarm is declared. Similarly, an omission occurs when no spike is detected within 0.5 ms of the true arrival time.

The performance of PMD was compared against the performances of the traditional amplitude thresholding detection (ATD) and CWD [30] methods. In addition, detections based on several state-of-the-art statistical techniques, namely a universal global threshold (UGT) [51], the false discovery rate (FDR) [52], and Stein’s unbiased risk estimate (SURE) [53], were also tested. The details of the CWD method can be found in [30]. UGT, FDR, and SURE represent wavelet-based techniques and are implemented similarly. Briefly, the signal features were extracted by applying CWT at relevant scales (see Section II-A), and for each scale a threshold was estimated based on the statistical properties of the wavelet coefficients and by optimizing (or nearly optimizing), a suitably defined objective function. The objective functions are: the Bayes risk in CWD, the mean-square risk in UGT detection, the false alarm

rate in FDR detection, and the estimate of the quadratic risk in SURE detection. The locations of the wavelet coefficients whose absolute values exceed the threshold represent the AP arrival time candidates. The candidate occurrence times are then processed (see Section II-D) to obtain a unique arrival time per detected AP. While statistically inferior, the widely used ATD method was evaluated for reference. In this case, arrival time candidates are simply defined as time samples whose absolute values cross a threshold. Note that this approach yields “double” thresholding and is necessary for truly unsupervised applications, since it is not known in advance whether the dominant phase of APs will be positive or negative [30]. The arrival time candidates were then processed in the same way as with the wavelet-based techniques.

1) *Receiver Operating Characteristics*: The performances of these methods were assessed by constructing receiver operating characteristic (ROC) curves, that is by plotting the probability of correct detection (P_{CD}) versus the probability of false alarm (P_{FA}) (see Fig. 3). For wavelet-based techniques, including PMD, these results are based on two relevant scales ($J = 2$). Similar results were obtained with 3-D features. For each SNR-FR combination, P_{FA} and P_{CD} represent the averages over 300 trials.

TABLE I
BIAS (FIRST ROW) AND STANDARD DEVIATION (SECOND ROW) OF THE ARRIVAL TIME ERROR ESTIMATED BY THE DIFFERENT METHODS OVER 300 TRIALS AT SNR = 3.5 AND FR = 40 HZ. THE LAST ROW SHOWS THE AVERAGE CPU TIME PER TRIAL

Method	ATD($m = 3.5$)	UGT	FDR($p = 0.001$)	CWD($r = 1$)	PMD	SURE
Mean error (ms)	0.032	0.048	0.001	0.039	0.060	-0.035
Std (ms)	0.177	0.155	0.200	0.177	0.187	0.215
CPU time per trial (sec)	0.89	1.51	1.97	3.79	26.16	1.77

For the three unsupervised techniques (PMD, UGT, and SURE), the performances are represented by a single point. The SURE detection had extremely high false alarm rates (occasionally $> 90\%$), and the corresponding ROCs have been omitted in Fig. 3 (to keep plots reasonably scaled). For the unsupervised techniques, the performances are given as a function of a detection sensitivity parameter [the false alarm-to-omission cost ratio r , (CWD), the p-value (FDR), and the multiple of the estimated standard deviation $\hat{\sigma}$ of noise (ATD)]. Intermediate performances are indicated by interpolating lines: (—) for CWD, (—) for ATD, and (—) for FDR detection. Equal cost of false alarm and omission ($r = 1$) was suggested in [30] for *unsupervised* applications of CWD, and the corresponding performance is marked by ■. The performance of FDR detection was evaluated for the following p-values: 10^{-1} , 10^{-3} , 10^{-5} , 10^{-7} (the points in Fig. 3 correspond to the two largest p-values). Finally, the performance of ATD was evaluated by varying the amplitude threshold $T_a = m\hat{\sigma}$, where $m = [3.5, 4.0, \dots, 5.5]$.

Based on Fig. 3, we conclude that PMD compares favorably to the supervised methods (CWD, FDR, and ATD) for various SNR-FR scenarios. The advantage is best seen in the bottom left panel of Fig. 3, where the PMD method achieves P_{CD} of 87.15% at the false alarm rate of 11.38% (marked by dotted lines). The correct detection rates of the CWD, FDR, and ATD methods at this level of P_{FA} are 82.79%, 81.43%, and 53.73%, respectively, and are inferior to that of PMD. Similarly, at the fixed correct detection rate of 87.15%, the PMD method achieves P_{FA} of 11.38%, while those of the CWD and FDR methods are 20.48% and 27.17%, respectively. [The false alarm rate of ATD at this level of P_{CD} has not been calculated, although it is clearly higher than those of the other techniques (see Fig. 3).] Since these observations are consistent for various SNR-FR combinations, we conclude that for a fixed P_{FA} , PMD attains higher P_{CD} than do the ATD, CWD and FDR methods. Likewise, at a fixed P_{CD} , PMD yields the lowest false alarm rates. Furthermore, the advantage of PMD appears more convincing at low SNRs, a situation commonly found in everyday recording practice. At high SNRs, the performances of these techniques are more comparable (with the exception of ATD). The comparison to the unsupervised techniques (UGT and SURE) as well as the unsupervised version of CWD is less straightforward and depends largely on the cost associated with the type I (false alarm) and type II (omission) errors. Nevertheless, the usefulness of the SURE method is severely limited by excessively large false alarm rates (averages over SNRs are: 93%, 84% and 67% for FRs of 10, 20, and 40 Hz, respectively). On the other hand, the UGT technique is quite conservative, resulting in a very low P_{FA} , and a relatively low P_{CD} , especially

at low SNRs. Finally, the performances of unsupervised CWD and PMD are somewhat similar, although the PMD method is more powerful (higher P_{CD}) and has a higher P_{FA} .

2) *Bias and Variance of Arrival Time Estimates*: An additional measure of the performance is provided by the accuracy of the estimated arrival times determined with each of the detection methods. Table I shows the mean and standard deviation of the error, defined as the difference between the true and estimated arrival times. The error is calculated over 300 trials at SNR = 3.5 and FR = 40 Hz. Similar results were observed for other SNR-FR combinations, but were omitted in the interest of space. The mean error quantifies the jitter, or bias, in the estimation of the arrival times, whereas its standard deviation provides a measure of the dispersion around the mean. It follows from Table I that all the methods have similar jitters, typically on the order of the sampling period of the data $T_s = 0.05$ ms. Likewise, the dispersion values are very consistent and typically range between $3 T_s$ and $4 T_s$. It can also be seen that the UGT and FDR methods trade off the bias and variance differently, and these results are in agreement with the literature on wavelet denoising [54].

3) *False Alarms Analysis*: We also analyzed the characteristic shapes of the false alarms for the ATD and PMD methods. The falsely detected APs were extracted from 50 trials, with FR = 40 Hz and SNR = 4.5, and classified in a 2-D feature space using principal components. Clusters were visually identified, and APs belonging to the same cluster were averaged to obtain templates, which represent the characteristic false alarm shapes. This procedure yielded three and four templates for the ATD and PMD method, respectively. Further analysis showed that the characteristic false alarm shapes of the ATD method were single-peaked monophasic waveforms, most likely due to random voltage fluctuations. On the other hand, the templates of the PMD method consisted of bi-phasic waveforms with a clear resemblance to APs. We hypothesize that this is a consequence of using the recorded noise to model neural data (see Section II-A-2). While the recorded noise contains no visually identifiable APs, it likely contains low amplitude APs, perhaps due to the activities of distant neurons in the background. This conjecture was confirmed by performing the above analysis with an autoregressive moving average (ARMA) noise model, which caused a considerable drop in the false alarm rate of the PMD method, whereas the ATD technique was rather insensitive to the change. Therefore, two conclusions can be drawn from this analysis. First, the majority of false alarms of the PMD method represent low-amplitude APs originating from distant neurons, and may be potentially useful for further analysis. Similar findings were reported in [30]. Second, ARMA noise models do not provide realistically adverse detection conditions, even if their

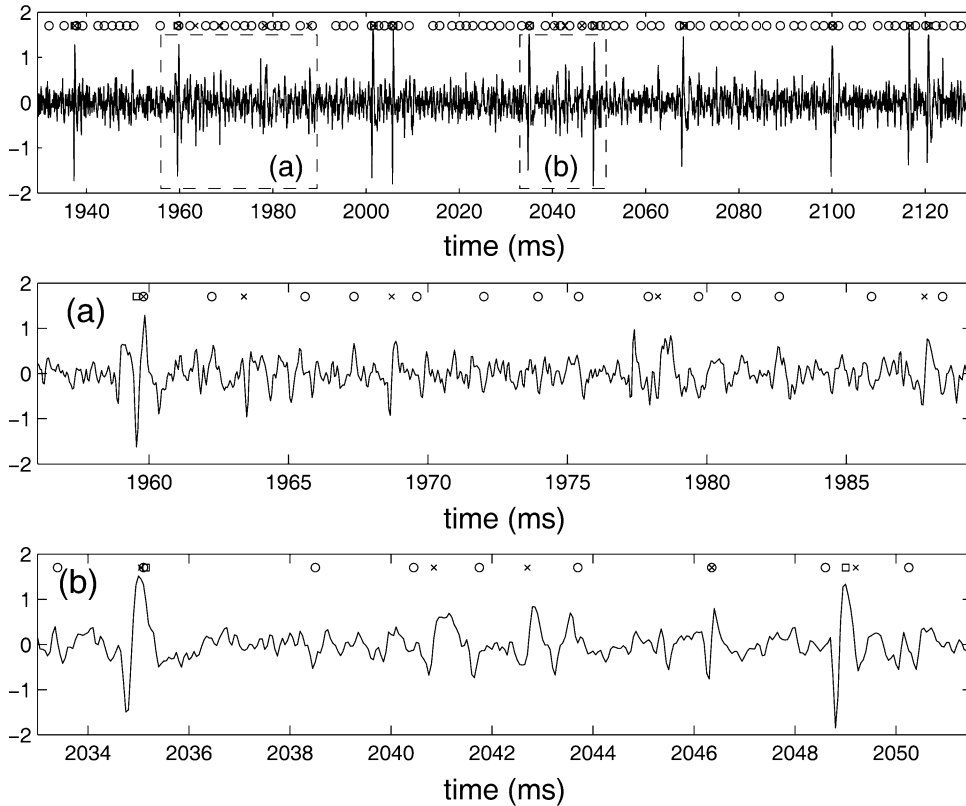


Fig. 4. Spikes detected from real data with various unsupervised methods. The top panel shows a segment of some 200 ms of data. The symbols above the signal indicate the estimated arrival times of APs detected by the unsupervised methods PMD (\times), UGT (\square), and SURE (\circ). The zoomed-in versions of the segments contained in the dashed boxes (a) and (b) are shown in the middle and bottom panels, respectively.

power spectra are well-matched to recorded data. Consequently, the use ARMA noise for evaluation of AP detection schemes (e.g., [55] and [56]) may yield false alarm rates that are optimistically biased.

4) *Performance With Real Data:* Finally, we tested the performance of the PDM method and other unsupervised methods (UGT and SURE) on actual extracellular recordings. The top panel of Fig. 4 shows real data recorded under medium SNR conditions. The shapes of APs detected by each method can be seen in the middle and bottom panels of Fig. 4. Clearly, the UGT method only detects high-amplitude APs and its performance is very conservative in terms of the false alarm rate, as was predicted based on the simulation results in Section III-B1. On the other hand, many of APs detected by the SURE method do not have a realistic AP shape, and are likely to represent false alarms (consistent with Section III-B1). Finally, the events detected by the PMD correspond to biphasic waveforms with a duration similar to the expected duration of an AP.

While the analysis with real data cannot provide quantitative results on the performance of the method, its findings are very consistent with the results based on Monte Carlo simulations. In particular, the PDM method compares favorably to the other unsupervised techniques.

IV. DISCUSSION

The results in Fig. 3 are consistent with several previous studies. Under high SNR, the choice of a detection method

becomes somewhat irrelevant, as the performance of any reasonable detector would provide good results. This is perhaps the main reason why ATD is still widely used in experimental studies, which are strongly biased toward high SNRs (experimentalists often report only their best data). Under low-SNR conditions, improved performances are typically achieved with more sophisticated detection techniques [17], [30]. The threshold in the SURE method is unrealistically low [29], thus the power of this technique is hindered by unacceptably high false alarm rates. On the contrary, the threshold set by the UGT method is too conservative, and yields performances with virtually no false alarms (see Fig. 3). Similar results have been reported in [30], where the UGT method was used to estimate the prior probabilities of signal and noise hypotheses.

The advantage of the PMD method over the other wavelet-based detection techniques stems from its superior modeling of the data in the feature domain. Fig. 5 shows detection boundaries according to CWD, UGT, and PMD. The detection boundaries of the FDR and SURE methods are similar and have been omitted in the interest of clarity. Clearly, the quadratic curve contour of PMD better separates the signal and noise, than do the rectangular contours of the other techniques. Furthermore, PMD takes into account the joint statistics of the coefficients across scales, while other techniques treat scales independently. We also note that data whitening, normally used to remove correlations, cannot be applied in a straightforward manner to CWD, UGT, and SURE techniques. Essentially, the scales are “mixed” linearly through the whitening transform

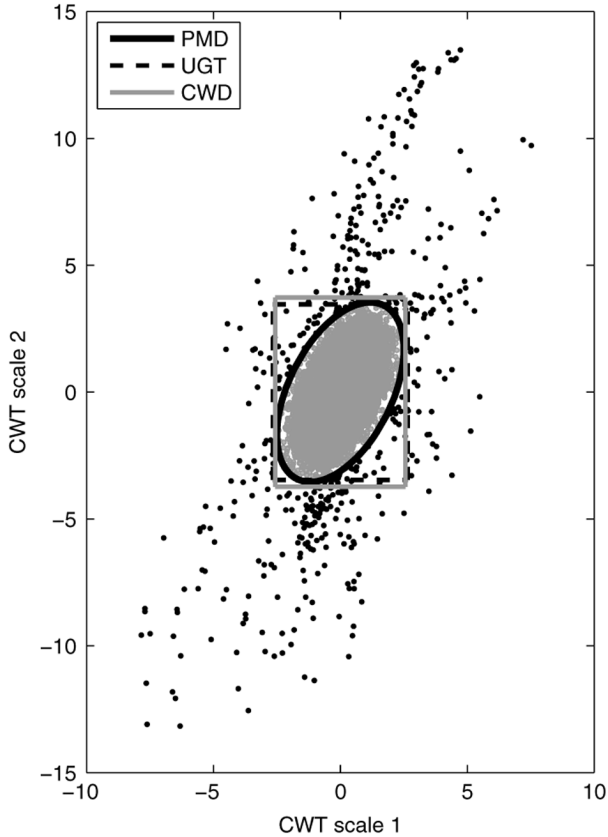


Fig. 5. Separation of signal and noise using different techniques: CWD ($r = 1$), PMD and UGT.

and it is unclear how the thresholds would be estimated. For the PMD method, whitening caused slight improvements, primarily due to better estimate of the uniform component in the model \mathcal{M}_2 .

We proceed by noting that our technique can be adopted to supervised applications. No attempt was made to optimize the choice of relevant scales throughout this study, which would amount to supervision. Rather, a generic choice of scales corresponding to AP duration of 0.5 and 1.5 ms was used. As discussed in Section II-A-2, these scales are biophysically relevant and are expected to work in a variety of experimental set-ups. For supervised applications, the relevant scales could be further optimized, and would consequently yield superior detection results. Moreover, the separation of signal and noise could be parametrized. Analogous to Section II-C, more or less evidence may be required to assign a feature, \mathbf{x}_i , to the signal or noise components, i.e.,

$$\mathbf{x}_i \in \begin{cases} \text{signal,} & \text{if } \tau_1^* \mathcal{U}(\mathbf{x}_i, V^*) > \alpha \tau_2^* \mathcal{G}(\mathbf{x}_i, \theta_2^*) \\ \text{noise,} & \text{otherwise} \end{cases}$$

where α is a (sensitivity) parameter. Choosing the optimal value of α amounts to trading off P_{FA} and P_{CD} and requires a supervision. The PMD technique developed in this article can be viewed as a special case of the above procedure, where instead of searching for the “right” α , its value is generically set to 1 (see Section II-C).

The main weakness of the PMD technique lies in its computational cost. The method is more computationally expensive than

the other techniques, mainly due to the use of the EM algorithm in the estimation of the probabilistic mixture model, \mathcal{M}_2 . The bottom row in Table I shows the average CPU times per trial in seconds, required by each of the detection methods using an Intel Core 2 Duo CPU4300@1.80 GHz processor. Despite being more computationally demanding, the PMD method is computationally affordable for any modern computer (around 2 s of processing for each second of data) while rendering a higher detection power.

V. CONCLUSION

By combining advanced signal processing and statistical tools, we have devised an AP detection algorithm that is suitable for unsupervised applications. The algorithm is adaptive in nature and provides a robust performance over a range of SNRs and FRs. At the core of our algorithm is the extraction of features of neural signals by means of the continuous wavelet transform, followed by the estimation of a probabilistic model in the feature domain. While our method does not strictly require the use of wavelet features, the choice of a suitable wavelet basis enables a sparse representation of APs, which consequently appear as outliers embedded in noise. The presence of APs is then cast as a simple model selection process, where the parameters of the model are estimated using the ML principle. Our method compares favorably to other unsupervised wavelet-based techniques, which are either overly conservative (UGT) or overly liberal (SURE). In comparison to supervised wavelet-based techniques (CWD and FDR), our method provides an advantage in performance, while being completely free of any detection sensitivity parameter. Failure mode analysis of our method shows that the majority of false alarms are low-amplitude APs originating from distant neurons, which could be potentially used in further analyses. Our method will be useful for many applications where human involvement is impractical, such as managing a large number of independently movable microelectrodes.

APPENDIX

Since the model \mathcal{M}_2 is a mixture of two components, the ML parameters τ^* and θ^* cannot be found analytically.⁶ The idea behind the EM algorithm is to maximize the likelihood

$$\mathcal{L}(\mathcal{X}|V^*, \tau, \theta) \triangleq \prod_{i=0}^{N-1} f(\mathbf{x}_i|V^*, \tau, \theta)$$

by introducing a latent variable \mathbf{z}_i , and so-called complete data $y_i \triangleq (\mathbf{x}_i, \mathbf{z}_i)$. The algorithm then proceeds by applying a two-step procedure to the log-likelihood⁷ of the complete data

$$\log \mathcal{L}(\mathcal{X}, \mathcal{Z}|\tau, \theta) \triangleq \sum_{i=0}^{N-1} \log f(\mathbf{x}_i, \mathbf{z}_i|\tau, \theta) \quad (9)$$

⁶An exception is the parameter V^* , calculated as $V^* = \prod_{j=1}^J M_j$, where $M_j = \max_{0 \leq i \leq N-1} |\mathbf{x}_i(j)|$, and $\mathbf{x}_i(j)$ is the j th component of the feature vector \mathbf{x}_i .

⁷Since log is a monotone function, maximizing the logarithm of a function is equivalent to maximizing the function itself.

where V^* has been dropped because it is a known quantity. While this appears to further complicate calculations, it actually simplifies the problem substantially. By iterating between the E-step

$$Q(\tau, \theta) = \mathbb{E}_{\mathcal{Z}|\mathcal{X}, \hat{\tau}, \hat{\theta}}[\log \mathcal{L}(\mathcal{X}, \mathcal{Z}|\tau, \theta) | \mathcal{X}, \hat{\tau}, \hat{\theta}] \quad (4)$$

where \mathbb{E} is the expectation operator, and the M-step

$$\{\hat{\tau}, \hat{\theta}\} = \arg \max_{\tau, \theta} Q(\tau, \theta)$$

where $\hat{\tau}$ and $\hat{\theta}$ are the current estimates of τ and θ , the (local) maximum $\{\tau^*, \theta^*\}$ of the original log-likelihood, $\log \mathcal{L}(\mathcal{X}|\tau, \theta)$, is approached [48]. By substituting (3) into (4), and after noting that \mathcal{Z} is a sequence of independent random variables, we have

$$\begin{aligned} Q(\tau, \theta) &= \sum_{i=0}^{N-1} \mathbb{E}_{\mathbf{z}_i|\mathbf{x}_i, \hat{\tau}, \hat{\theta}} \\ &\quad [\log f(\mathbf{x}_i, \mathbf{z}_i|\tau, \theta) | \mathbf{x}_i, \hat{\tau}, \hat{\theta}] \\ &= \sum_{i=0}^{N-1} \sum_{\text{all } \mathbf{z}_i} \log f(\mathbf{x}_i, \mathbf{z}_i|\tau, \theta) \\ &\quad \times P(\mathbf{z}_i|\mathbf{x}_i, \hat{\tau}, \hat{\theta}) \end{aligned} \quad (5)$$

where we have assumed that \mathcal{Z} is a discrete random variable.

To proceed, the latent variable \mathcal{Z} needs to be specified. While its definition is not unique, \mathcal{Z} should indicate the class membership of each feature \mathbf{x}_i . Adopting the suggestion of Fraley and Raftery [40] to our two-component model, \mathcal{M}_2 , results in

$$\mathbf{z}_i \triangleq \begin{cases} [1, 0], & \text{if } \mathbf{x}_i \in \mathcal{U} \quad (\text{signal}) \\ [0, 1], & \text{if } \mathbf{x}_i \in \mathcal{G} \quad (\text{noise}) \end{cases}$$

where \mathbf{z}_i takes these values with probability τ_1 and τ_2 , respectively. An alternative definition of \mathbf{z}_i is given in [57]. The joint PDF in (5) can be factored out as $f(\mathbf{x}_i|\mathbf{z}_i, \tau, \theta)P(\mathbf{z}_i|\tau, \theta)$, which based on the definition of \mathbf{z}_i yields

$$f(\mathbf{x}_i, \mathbf{z}_i|\tau, \theta) = \begin{cases} \tau_1 \mathcal{U}(\mathbf{x}_i), & \text{if } \mathbf{z}_i = [1, 0] \\ \tau_2 \mathcal{G}(\mathbf{x}_i, \theta), & \text{if } \mathbf{z}_i = [0, 1]. \end{cases}$$

To complete the calculation of (5), the posterior probability needs to be evaluated, which follows readily from Bayes' rule

$$P(\mathbf{z}_i|\mathbf{x}_i, \hat{\tau}, \hat{\theta}) = \frac{f(\mathbf{x}_i|\mathbf{z}_i, \hat{\tau}, \hat{\theta})P(\mathbf{z}_i|\hat{\tau}, \hat{\theta})}{\sum_{\text{all } \mathbf{z}_i} f(\mathbf{x}_i|\mathbf{z}_i, \hat{\tau}, \hat{\theta})P(\mathbf{z}_i|\hat{\tau}, \hat{\theta})}. \quad (6)$$

The individual terms in (6) are found as

$$f(\mathbf{x}_i|\mathbf{z}_i, \hat{\tau}, \hat{\theta})P(\mathbf{z}_i|\hat{\tau}, \hat{\theta}) = \begin{cases} \hat{\tau}_1 \mathcal{U}(\mathbf{x}_i), & \text{if } \mathbf{z}_i = [1, 0] \\ \hat{\tau}_2 \mathcal{G}(\mathbf{x}_i, \hat{\theta}), & \text{if } \mathbf{z}_i = [0, 1]. \end{cases}$$

Once computed from the E-step, $Q(\tau, \theta)$ can be maximized in a straightforward fashion, although derivations are rather involved, and will not be treated here. Instead, the interested reader is referred to [57]. Fortunately, final results are quite simple and are summarized below ($k = 1, 2$):

$$N_k \triangleq \sum_{i=0}^{N-1} \hat{z}_{ik} \quad (7)$$

$$\hat{\tau}_k \leftarrow \frac{N_k}{N} \quad (8)$$

$$\hat{\mu} \leftarrow \frac{1}{N_2} \sum_{i=0}^{N-1} \hat{z}_{i2} \mathbf{x}_i \quad (9)$$

$$\hat{\Sigma} \leftarrow \frac{1}{N_2} \sum_{i=0}^{N-1} \hat{z}_{i2} (\mathbf{x}_i - \hat{\mu})(\mathbf{x}_i - \hat{\mu})^T \quad (10)$$

where

$$\begin{aligned} \hat{z}_{i1} &\triangleq \frac{P(\mathbf{z}_i = [1, 0]|\mathbf{x}_i, \hat{\tau}, \hat{\theta})}{\hat{\tau}_1 \mathcal{U}(\mathbf{x}_i) + \hat{\tau}_2 \mathcal{G}(\mathbf{x}_i, \hat{\theta})} \\ &= \frac{P(\mathbf{z}_i = [1, 0]|\mathbf{x}_i, \hat{\tau}, \hat{\theta})}{\hat{\tau}_1 \mathcal{U}(\mathbf{x}_i) + \hat{\tau}_2 \mathcal{G}(\mathbf{x}_i, \hat{\theta})} \end{aligned} \quad (11)$$

$$\begin{aligned} \hat{z}_{i2} &\triangleq \frac{P(\mathbf{z}_i = [0, 1]|\mathbf{x}_i, \hat{\tau}, \hat{\theta})}{\hat{\tau}_1 \mathcal{U}(\mathbf{x}_i) + \hat{\tau}_2 \mathcal{G}(\mathbf{x}_i, \hat{\theta})} \\ &= \frac{P(\mathbf{z}_i = [0, 1]|\mathbf{x}_i, \hat{\tau}, \hat{\theta})}{\hat{\tau}_1 \mathcal{U}(\mathbf{x}_i) + \hat{\tau}_2 \mathcal{G}(\mathbf{x}_i, \hat{\theta})}. \end{aligned} \quad (12)$$

Thus, for practical purposes the algorithm reduces to updating \hat{z}_{i1} and \hat{z}_{i2} in (11) and (12) (E-step), followed by updating the parameters in (7), (8), (9), and (10) (M-step). We initialized the above scheme by partitioning data into two clusters: $G \triangleq \{\mathbf{x}_i : \mathcal{D}(\mathbf{x}_i) \leq 3.5\}$ and $U \triangleq \mathcal{X} \setminus G$, where $\mathcal{D}(\mathbf{x}) \triangleq (\mathbf{x}^T \hat{\Sigma}_m^{-1} \mathbf{x})^{1/2}$ stands for the Mahalanobis distance, and $\hat{\Sigma}_m$ is the sample covariance of \mathcal{X} . From this, we obtain the initial value of $\hat{\theta} = \{\hat{\mu}, \hat{\Sigma}\}$, where $\hat{\mu}$ and $\hat{\Sigma}$ are the sample mean and sample covariance of G , respectively. Similarly, $\hat{\tau}_1 = \|U\|/N$ and $\hat{\tau}_2 = \|G\|/N$, where $\|\cdot\|$ stands for the size of a set.

REFERENCES

- [1] E. D. Adrian, "The impulses produced by sensory nerve endings," *J. Physiol. (Lond.)*, vol. 61, pp. 49–72, 1926.
- [2] A. P. Georgopoulos, J. F. Kalaska, R. Caminiti, and J. Massey, "On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex," *J. Neurosci.*, vol. 2, no. 11, pp. 1527–1537, 1982.
- [3] W. Bialek and A. Zee, "Coding and computation with neural spike trains," *J. Stat. Phys.*, vol. 59, no. 1–2, pp. 103–115, 1990.
- [4] D. Hubel, "Tungsten microelectrode for recording from single units," *Science*, vol. 125, pp. 549–550, 1957.
- [5] J. Green, "A simple microelectrode for recording from the central nervous system," *Nature*, vol. 182, p. 962, 1958.
- [6] D. A. Robinson, "The electrical properties of metal microelectrodes," *Proc. IEEE*, vol. 56, no. 6, pp. 1065–1071, 1968.
- [7] D. Humphrey, *Electrophysiological Techniques*. Atlanta, GA: Soc. for Neuroscience, 1979.
- [8] R. Lemon, *Methods for Neuronal Recording in Conscious Animal*. New York: Wiley, 1984, ch. 2, pp. 17–38.
- [9] J. Welsh and C. Schwarz, M. Nicoletis, Ed., "multielectrode recording from the cerebellum," in *Methods for Neural Ensemble Recordings*. Boca Raton, FL: CRC Press, 1999, ch. 5, pp. 79–100.
- [10] K. Frank and M. C. Becker, W. L. Natsuk, Ed., "Microelectrodes for recording and stimulation," in *Physical Techniques in Biological Research*. New York: Academic, 1964, vol. V, ch. 2, pp. 22–87.
- [11] M. S. Fee, P. P. Mitra, and D. Kleinfeld, "Variability of extracellular spike waveforms of cortical neurons," *J. Neurophysiol.*, vol. 69, pp. 175–188, 1996.
- [12] J. A. White, J. T. Rubinstein, and A. R. Kay, "Channel noise in neurons," *Trends Neurosci.*, vol. 23, pp. 131–137, 2000.
- [13] G. L. Gerstein and W. A. Clark, "Simultaneous studies of firing patterns in several neurons," *Science*, vol. 143, pp. 1325–1327, 1964.
- [14] M. Abeles and M. H. Goldstein, "Multispikes train analysis," *Proc. IEEE*, vol. 65, pp. 762–773, 1977.
- [15] X. Yang and S. A. Shamma, "A totally automated system for the detection and classification of neural spikes," *IEEE Trans. Biomed Eng.*, vol. 35, no. 10, pp. 806–816, Oct. 1988.

- [16] S. Yamada, H. Kage, M. Nakashima, S. Shiono, and M. Maeda, "Data processing for multi-channel optical recording: action potential detection by neural network," *J. Neurosci. Methods*, vol. 43, no. 1, pp. 23–33, 1992.
- [17] I. Bankman, K. Johnson, and W. Schneider, "Optimal detection, classification, and superposition resolution in neural waveform recordings," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 8, pp. 836–841, Aug. 1993.
- [18] S. N. Gozani and J. P. Miller, "Optimal discrimination and classification of neuronal action potential waveforms from multiunit, multi-channel recordings using software-based linear filters," *IEEE Trans. Biomed. Eng.*, vol. 41, no. 4, pp. 358–372, Apr. 1994.
- [19] K. D. Wise, "Micromachined interfaces to the cellular world," *Sensor. Mater.*, vol. 10, no. 6, pp. 385–395, 1998.
- [20] P. Norlin, M. Kindlundh, A. Mouroux, K. Yoshida, and U. G. Hofmann, "A 32-site neural recording probe fabricated by DRIE of SOI substrates," *J. Micromech. Microeng.*, vol. 14, no. 4, pp. 414–419, 2002.
- [21] G. Buzsaki, "Large-scale recording of neuronal ensembles," *Nature Neurosci.*, vol. 7, pp. 446–451, 2004.
- [22] J. G. Cham, E. A. Branchaud, Z. Nenadic, B. Greger, R. A. Andersen, and J. W. Burdick, "Semi-chronic motorized microdrive and control algorithm for autonomously isolating and maintaining optimal extracellular action potentials," *J. Neurophysiol.*, vol. 93, no. 1, pp. 570–579, 2005.
- [23] Z. Nenadic and J. W. Burdick, "A control algorithm for autonomous optimization of extracellular recordings," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 5, pp. 941–955, May 2006.
- [24] M. S. Lewicki, "Bayesian modeling and classification of neural signals," *Neural Comput.*, vol. 6, pp. 1005–1030, 1994.
- [25] M. Sahani, "Latent Variable Models for Neural Data Analysis," Ph.D. dissertation, California Inst. Technology, Pasadena, CA, 1999.
- [26] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998.
- [27] H. L. V. Trees, *Detection, Estimation and Modulation Theory*. New York: Appleton & Lange, 1991.
- [28] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
- [29] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1999.
- [30] Z. Nenadic and J. W. Burdick, "Spike detection using the continuous wavelet transform," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 1, pp. 74–87, Jan. 2005.
- [31] C. Gold, D. A. Henze, C. Koch, and G. Buzsaki, "On the origin of the extracellular action potential waveform: A modeling study," *J. Neurophysiol.*, vol. 95, pp. 3113–3128, 2006.
- [32] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
- [33] K. G. Oweiss, "Multiresolution Analysis of Multichannel Neural Recordings in the Context of Signal Detection, Estimation, Classification and Noise Suppression," Ph.D. dissertation, Univ. Michigan, Ann Arbor, 2002.
- [34] M. Frisch and H. Masser, "The use of the wavelet transform in the detection of an unknown transient signal," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 892–897, 1992.
- [35] N. Erdol, A. Akansu and M. Medley, Eds., "Multiscale detection," in *Wavelets, Subband and Block Transforms in Communications and Media*. Boston, MA: Kluwer Academic, 1999, ch. 6, pp. 183–205.
- [36] Z. Nenadic and J. Burdick, "Robust unsupervised detection of action potentials using the wavelet transform," in *Soc. Neurosci. Abstr.*, 2003, vol. 33, p. 279.1.
- [37] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," Yale Univ., New Haven, CT, and Stanford Univ., Stanford, CA, Tech. Rep., 1995 [Online]. Available: <http://citeseer.ist.psu.edu/80329.html>
- [38] G. McLachlan and K. Basford, *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker, 1988.
- [39] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [40] C. Fraley and A. E. Raftery, "Model-based clustering, discriminant analysis, and density estimation," *J. Amer. Stat. Assoc.*, vol. 97, no. 458, pp. 611–631, 2002.
- [41] C. Fraley and A. E. Raftery, "How many clusters? Which clustering method? answers via model-based cluster analysis," *Comput. J.*, vol. 41, pp. 578–588, 1998.
- [42] E. Jaynes, G. L. Bretthorst, Ed., *Probability Theory: The Logic of Science*. New York: Cambridge Univ. Press, 2003.
- [43] R. E. Kass and A. E. Raftery, "Bayes factors," *J. Amer. Stat. Assoc.*, vol. 90, pp. 773–795, 1995.
- [44] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, 1978.
- [45] M. Stone, "Cross-validated choice and assesment of statistical predictions," *J. Roy. Statist. Soc. B*, vol. 36, pp. 111–147, 1974.
- [46] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, pp. 716–723, 1974.
- [47] P. Cheeseman and J. Stutz, U. Fayyad, G. Paitesky-Shapiro, P. Smyth, and R. Uthurusamy, Eds., "Bayesian classification (autoclass): Theory and results," in *Advances in Knowledge Discovery and Data Mining*. Menlo Park, CA: AAAI Press, 1996, pp. 153–180.
- [48] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J. Roy. Stat. Soc. B Met.*, vol. 39, no. 1, pp. 1–38, 1977.
- [49] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, "Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering," *Neural Comput.*, vol. 16, no. 8, pp. 1661–1687, 1994.
- [50] K. H. Kim and S. J. Kim, "Neural spike sorting under nearly 0-db signal-to-noise ratio using nonlinear energy operator and artificial neural network classifier," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 10, pp. 1406–1411, Oct. 2000.
- [51] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
- [52] F. Abramovich and Y. Benjamini, "Adaptive thresholding of wavelet coefficients," *Comput. Stat. Data An.*, vol. 22, no. 4, pp. 351–361, 1996.
- [53] C. M. Stein, "Estimation of the mean of the multivariate normal distribution," *Ann. Stat.*, vol. 9, no. 6, pp. 1135–1151, 1981.
- [54] F. Abramovich and Y. Benjamini, "Adaptive thresholding of wavelet coefficients," *Comput. Stat. Data Anal.*, vol. 22, pp. 351–361, 1996.
- [55] K. H. Kim and S. J. Kim, "Neural spike sorting under nearly 0-db signal-to-noise ratio using nonlinear energy operator and artificial neural-network classifier," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 10, pp. 1406–1411, Oct. 2000.
- [56] K. H. Kim and S. J. Kim, "A wavelet-based method for action potential detection from extracellular neural signal recording with low signal-to-noise ratio," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 8, pp. 999–1011, 2003.
- [57] J. Bilmes, A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models Univ. of California, Berkeley, CA, Tech. Rep., Apr. 1998.



Raul Benitez (M'07) received the B.S. degree in physics from the University of Barcelona, Spain, in 1996, and the M.S. and Ph.D. degrees in computational physics from the Technical University of Catalonia, Barcelona, Spain, in 1999 and 2005, respectively.

From 2005 to 2006, he was a Postdoctoral Fellow with the Department of Biomedical Engineering at the University of California, Irvine. He is currently a Professor in the Automatic Control Department at the Technical University of Catalonia. His research

interests include analysis and control of nonlinear systems, biomedical signal processing and mathematical modeling of biological systems.



Zoran Nenadic (M'06) received the Diploma degree in control engineering from the University of Belgrade, Serbia, in 1995 and the M.S. and D.Sc. degrees in systems science and mathematics from Washington University, St. Louis, MO, in 1998 and 2001, respectively.

From 2001 to 2005, he was a Postdoctoral Fellow with the Division of Engineering and Applied Science at the California Institute of Technology, Pasadena. Since 2005, he has been with the Department of Biomedical Engineering, University of California, Irvine, where he is currently an Assistant Professor. His research interests are in the area of adaptive biomedical signal processing, control algorithms for biomedical devices, brain-machine interfaces, and modeling and analysis of biological neural networks.

Dr. Nenadic is a member of the Mathematical Association of America and the Society for Neuroscience.