ELSEVIER

Journal of Neuroscience Methods 117 (2002) 1–12

JOURNAL OF NEUROSCIENCE METHODS

www.elsevier.com/locate/jneumeth

A method for spike sorting and detection based on wavelet packets and Shannon's mutual information

Eyal Hulata, Ronen Segev, Eshel Ben-Jacob *

Raymond & Beverly Sackler Faculty of Exact Sciences, School of Physics and Astronomy, Tel-Aviv University, Tel-Aviv 69978, Israel

Received 1 November 2001; received in revised form 14 February 2002; accepted 15 February 2002

Abstract

Studying the dynamics of neural activity via electrical recording, relies on the ability to detect and sort neural spikes recorded from a number of neurons by the same electrode. We suggest the wavelet packets decomposition (WPD) as a tool to analyze neural spikes and extract their main features. The unique quality of the wavelet packets-adaptive coverage of both time and frequency domains using a set of localized packets, facilitate the task. The best basis algorithm utilizing the Shannon's information cost function and local discriminant basis (LDB) using mutual information are employed to select a few packets that are sufficient for both detection and sorting of spikes. The efficiency of the method is demonstrated on data recorded from in vitro 2D neural networks, placed on electrodes that read data from as many as five neurons. Comparison between our method and the widely used principal components method and a sorting technique based on the ordinary wavelet transform (WT) shows that our method is more efficient both in separating spikes from noise and in resolving overlapping spikes.

Keywords: Spike sorting; Wavelets; Wavelet packets; Shannon's information; Multi-electrodes

1. Introduction

Current electronics enables the recording of in vitro neural networks cell cultures, hippocampal slices, and in vivo tissues by simultaneous measurement of the activity of many neurons using multichannel electrode arrays (Jimbo et al., 1993; Stenger and Mckenna, 1994; Camepari et al., 1997; Fromherz et al., 1997; Egert et al., 1998; Maher et al., 1999). It is expected to improve our understanding of how real neural networks code, learn and store information.

Extracting useful information from these measurements relies on the ability to detect and sort the recorded neural spikes (Lewicki, 1998). Any method for detecting and sorting of spikes is comprised of two stages: (1) extracting the important features of the spikes and thus lowering the dimension of the parametric set representing the spikes. (2) Clustering of the parametric sets into groups, where each cluster relates to a different spike type.

The clustering problem is widely known and there are many methods to solve it. The performance of any clustering method relies on the original differences among the groups in the data. Hence, the efficiency of the feature extraction method in signifying those differences is crucial for the success of clustering. In this paper, we present a novel feature extraction method that efficiently lowers the dimension while signifying the differences among spike types.

Many powerful clustering methods have been introduced in spike sorting techniques (Lewicki, 1994, 1998; Rinberg et al., 1998; Zouridakis and Tam, 2000). However, we will show that the superior performance of our feature extraction technique enables us to successfully cluster the data using merely the k-means technique, which is one of the simplest methods known.

A new method to extract the features of recorded neural spikes is the wavelet transform (WT) (Unser and Aldroubi, 1996; Zouridakis and Tam, 1997; Yang and Shamma, 1998; Letelier and Wever, 2000). The neural spikes, being short voltage pulses, are localized func-

* Corresponding author. Tel.: +972-3-640-7845; fax: +972-3-642-2979.

E-mail address: eshel@venus.tau.ac.il (E. Ben-Jacob).

0165-0270/02/$ - see front matter
PII: S0165-0270(02)00032-8

tions whose natural analysis is via the WT (Unser and Aldroubi, 1996; Yang and Shamma, 1998), which is the Fourier transform analog for localized functions. Recent publications have shown good performance of the WT in detection and sorting of artificial data (Zouridakis and Tam, 1997; Letelier and Wever, 2000). Yet, the WT does not perform well in separating spikes that overlap or are even close in time.

Here we present a more efficient method based on WPD (Hulata et al., 2000), which is a generalization of the WT. The wavelet packets (Coifman et al., 1992; Daubechies, 1992; Mallat, 1998) are a large family of localized functions, where sets of packets can be adaptively selected to span signals of various characteristics. Algorithms such as best basis and local discriminant basis (LDB) (Coifman et al., 1992; Coifman and Wickerhauser, 1992; Saito, 1994; Mallat, 1998) are used to select sets of packets that enhance the features of the signal using a small number of significant coefficients. In the case of neural spikes, these packets form a compact representation, with minimal coverage of nearby spikes. This gives rise to detection and sorting of spikes even in cases of superimposed (overlapping) spikes. We present the efficiency of our method utilizing the selection of only nine packets. These packets were tested for many different spike waveforms on different electrodes in different experiments, and appear universal for recordings of this kind.

All numerical procedures were coded using MATLAB (www.mathworks.com) and the WAVELAB toolbox (Donoho et al., 2000).

2. Definition of the problem

The nature of the task is presented in Fig. 1. Fig. 1a presents a recording of a typical neural spike. Fig. 1b presents an example of recorded noise. It was recorded due to the strong peak amplitude. However, it is not a neural spike as the width of the event is less than 1 ms, and it lacks a refractory period. Fig. 1c presents a recording of a number of spikes from five different neurons that fired close in time.

The data acquisition is based on a hardware threshold detector that is adjusted to detect voltage amplitudes, which exceed the background noise by 3 folds (the background noise is estimated during non-firing periods). Whenever an event of amplitude above the threshold is detected, a window is recorded with the event at its center. In our case the length of the window is 20 ms.

Summarizing the main obstacles, a good spike sorting and detection method must carry out the following:

1) detection of spikes in the presence of noise, and distinction between neural spikes and noise events.
2) Sorting among different neurons read by the same electrode utilizing the differences in the spike waveforms.
3) Detection of close fired spikes, and identification of the origin in presence of superposition. Note that when superposition occurs (as in Fig. 1c), the spike's waveform is not complete and that complicates the identification.

3. Wavelet transform versus Fourier analysis

The Fourier transform of a signal $f(t)$ is generally defined by:

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t) \mathrm{e}^{-i\omega t}\, \mathrm{d}t \tag{1}$$

When $f(t)$ is measured over a finite time interval T, and sampled every time interval τ, the Fourier transform can also be viewed as a set of N filters (where $N = T/\tau$). The filters cover the frequency region $[-\pi/\tau, \pi/\tau]$, each with a bandwidth of $\Delta\omega = 2\pi/T$. The energy distribution of $f(t)$ over the frequency interval is $|\hat{f}(\omega)|$, and $\int_{-\pi/\tau}^{\pi/\tau} |\hat{f}(\omega)|^2 \mathrm{d}\omega$ is the total energy of $f(t)$. Note that $\hat{f}(\omega)$ does not contain any temporal information.

The WT was devised to maintain both the temporal and frequency properties of a signal. This analysis is based on a complete set of localized functions (named wavelets) that span both the time and frequency domains. The WT of $f(t)$ is thus a function of two parameters a and b, where a represents a frequency scale, and b indicates the time location of the wavelet (Daubechies, 1992; Mallat, 1998):

$$W(a,\ b) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{a}} \psi_{a,b}(t) \mathrm{d}t \quad \text{where}$$
$$\psi_{a,b}(t) = \psi\left(\frac{t-b}{a}\right) \tag{2}$$

$\psi(t)$ is called the mother wavelet and $\psi_{a,b}(t)$ is a scaled and translated wavelet. Note that $\psi_{a,b}(t)$ replaces the $\mathrm{e}^{-i\omega t}$ of Eq. (1). However, unlike the Fourier transform, there are numerous mother wavelets, and $\psi(t)$ is chosen according to the problem. For example, Fig. 2 presents the mother wavelet we use, the 3rd order Coiflet mother wavelet, and a number of scaled wavelets. The inverse WT is:

$$f(t) = \frac{1}{C_\psi} \int_0^{\infty} \int_{-\infty}^{\infty} W(a,\ b) \frac{1}{\sqrt{a}} \psi_{a,b}(t) \mathrm{d}b \frac{\mathrm{d}a}{a^2}$$
$$C_\psi = \int_0^{\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega}\, \mathrm{d}\omega \tag{3}$$

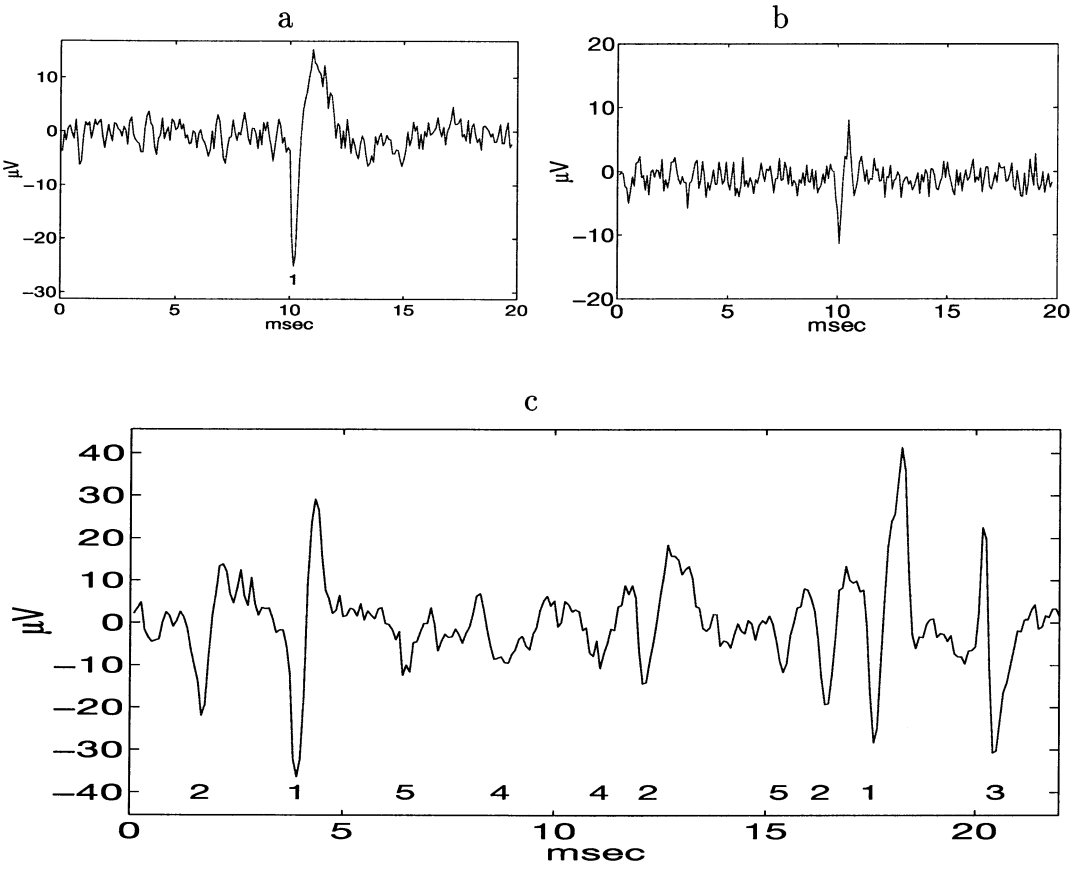

Fig. 1. Windows recorded from a single electrode. (a) A typical neural spike. (b) Recording of noise. (c) Recording from five different neurons (marked 1–5 beneath each detected and sorted spike). Note the very short durations between the spikes, and the poor signal to noise ratio of the spikes from neurons 4 and 5.

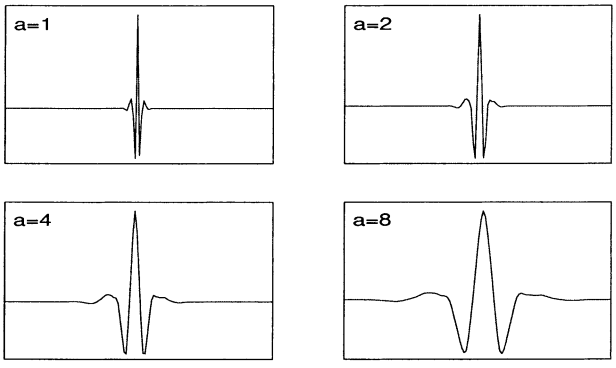

Fig. 2. Third order Coiflet mother wavelet and scaled wavelets. The scale a defines the temporal width of the wavelet (see Eq. (2)). Arbitrary units.

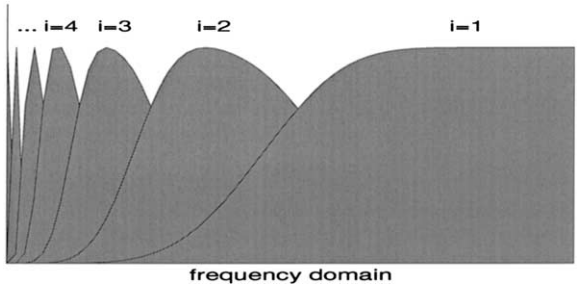

Fig. 3. The division of the frequency domain by the wavelet transform. Each scale $a = 2^i$, corresponds to a single bandpass filter of width $\pi/a\tau$.

For a signal $f(t)$ measured over time T and sampled at a time scale of τ, the WT can also be viewed as a discrete set of filters. In contrast to the Fourier transform, the set of filters of the WT are not equal in bandwidth. The filters vary between the narrowest filter $\Delta\omega_{\min} = \pi/T$ to the widest filter $\Delta\omega_{\max} = \pi/2\tau$, which is half the frequency domain. The width of the filter is determined by the value of the scale parameter a: $\Delta\omega = \pi/a\tau$. The discrete values of a are of the type $a = 2^i$ (where $i = 1, \ldots, L$ and $L = \log_2(N)$). Fig. 3 presents this set of filters. The filters are organized such that the widest filter covers the higher part of the frequency domain, and as i increases, both the bandwidth is narrower and the filter location is lower. The filters also divide the time domain. The uncertainty principal, $\Delta\omega\Delta t \geq \pi$, imposes that as the filter is narrower in bandwidth, the resolution in time is lowered. As a consequence, when the bandwidth is $\Delta\omega_{\min}$, $\Delta t = T$ and we lose all temporal information. As $\Delta\omega$ increases, time resolution is gained. The concept

of both time and frequency division is presented in Fig. 6b. In this figure, a neural spike is presented. It's Fourier transform is plotted in vertical, and the WT is shown as a 2D surface, which represents the time–frequency space. The surface is covered by rectangles, where the two dimensions represent the frequency and temporal resolutions.

Due to the localization property of the wavelets, the distribution of the signal's energy $|W(a, b)|^2$ is better concentrated than in the Fourier transform. A comparison between the Fourier and WTs of a neural spike is presented in Fig. 4. In contrast to the WT, the Fourier transform of the signal does not present any profound features.

The usage of the WT for analyzing neural spikes is not optimal. For large values of a (i.e. large scales), the time resolution of the wavelets are wider than the width of the spikes. Therefore, when recording overlapping spikes, the wavelets fail to separate them. Fig. 5 presents a window with three recorded spikes, and the reconstruction of the event using coefficients of the WT. As the wavelets are too wide, they simultaneously cover two spikes. Another example of the inadequate resolution is seen in Fig. 6a. The rectangles that bear the energy of the spike extend beyond the length of the spike both in time and frequency domains.

The time–frequency space can be divided using a more adaptive set of functions that better localize the spikes. That is implemented via the WPD. Fig. 6b presents the division of the time–frequency space according to the WPD. Note that the rectangles localize the spike better than in Fig. 6a. Fig. 5 presents the reconstruction using WPD of a window with multiple spikes. Note that the WPD is able to separate the spikes, not like the WT.

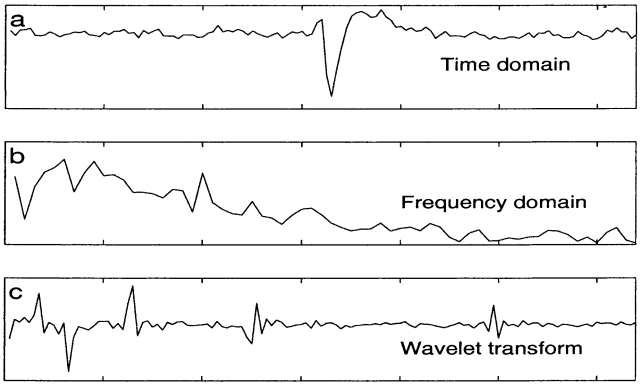

Fig. 4. Comparison between a neural spike in time domain (a), its Fourier transform (b) and wavelet transform (c). Note that in the Fourier domain (b) the energy is disturbed over a wide bandwidth, with no favorable frequencies. In the WT representation (c), the coefficients are lined according to their frequency range, from low to high frequencies. Note that the energy of the spike is concentrated using very few wavelets, whereas the noise is disturbed almost evenly over all the other wavelets.

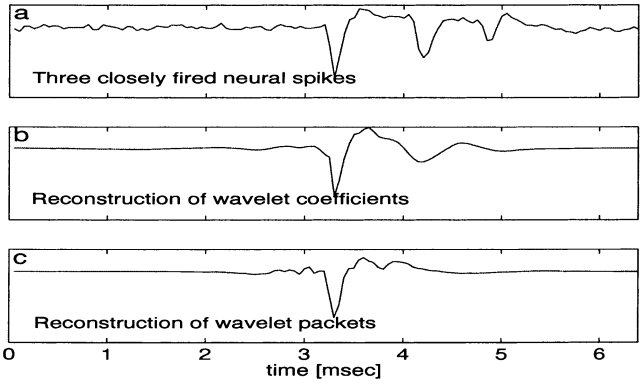

Fig. 5. Comparison between the representation of closely fired spikes using standard WT and WPD. (a) The spikes in time domain. (b) Reconstruction of the spike in the middle using the nine most energetic wavelets. (c) Reconstruction of the spike using nine selected wavelet packets. This quality of reconstruction and the seclusion of the spike demonstrate the efficiency of using optimized wavelet packets. Note that WT did not separate the spike from its consecutive since high scale wavelets are wide and cover more than a single spike. However, the width of the optimized selected packets enable the successful reconstruction of the analyzed spike alone.

4. Overview of wavelet packets decomposition

The WPD generalizes the filtering concept of the WT. In the WT, only one filter is defined for each scale a (with bandwidth $\Delta\omega = \pi/a\tau$), thus forming a set of filters each half in bandwidth than it's former. In contrast, the key characteristic of the WPD is that each bandwidth is used for forming a complete coverage of the frequency domain, thus creating multiple divisions of the domain. Fig. 7a illustrates this division. The width of the filters is related to the parameter i exactly as in the WT. Note that for each value of i, a series of filters with an equal bandwidth (numbered using the parameter j) cover the entire frequency domain. Thus, the wavelet packets constitute a family of functions characterized by three parameters: $\omega_{i,j,k}(t)$ where $i = 1, \ldots, \log_2(N)$ is the level of decomposition, $j = 0, \ldots, 2^i - 1$ designates the region that is covered on the frequency domain, and k indicates the temporal location of the packet (equivalent to b in the WT). Any selection of i and j indicates a frequency region (also called block) of width $\Delta\omega = \pi/2^i\tau$ covering the range $[j\Delta\omega, (j+1)\Delta\omega]$.

As seen in Fig. 7a, this form of filtering creates an over-complete division of the time–frequency space, as every region in both frequency and time domain is covered by more than one block. Any set of blocks that covers the frequency domain without overlapping, also covers the time domain (Fig. 7b presents an example of a possible set). A non-overlapping set of blocks is commonly referred to as a basis, as it spans the time–frequency space. Hence, when analyzing a signal $f(t)$, we would like to find a basis that optimally localizes the features of $f(t)$.

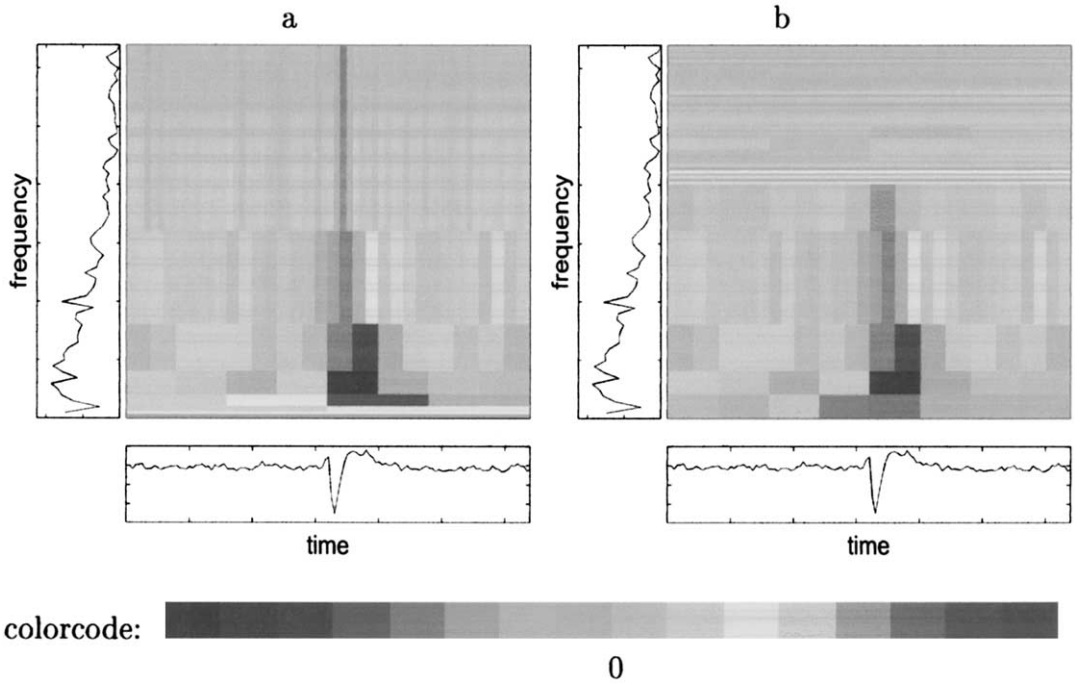

Fig. 6. Comparison between WT (a) and WPD (b) on a 2D time–frequency space. The time and frequency domains are presented as X and Y axes. (a) The 2D space is plotted via the division of the domains by the WT. The color code represents the distribution of energy over the different wavelets. Note that the features of the spike are stretched over the entire range of frequencies. (b) The 2D space is plotted via the division of the domains by the packets in our LDB. Note that the features of the spike are localized in an area bounded both in time and frequency, and in a better fashion than in the WT case.

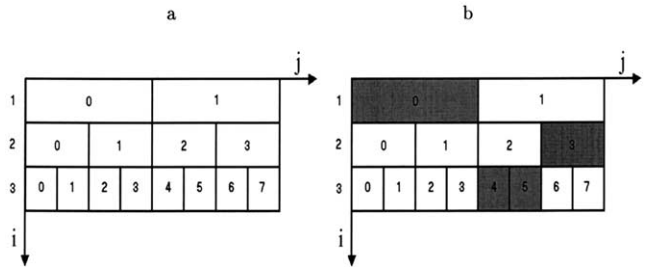

Fig. 7. (a) The division of the frequency domain by the wavelet packets decomposition: in each level i, the packets divide the domain into 2^i adjacent blocks, all of equal bandwidth. For example, the block $i=2, j=1$ covers the frequency region $[W/4, W/2]$, where $W=\pi/\tau$ is the total bandwidth. Each packet in this block is of width $\Delta t=4\tau$. (b) An example of a set of blocks that cover the frequency domain without overlapping. This set is referred to as a basis that spans the time–frequency space.

4.1. Best orthonormal basis

The best basis algorithm is a method to select a basis that spans the signal with the smallest number of significant coefficients. The idea is to find a way to favor either a block with parameters $[i, j]$ or the corresponding next level blocks labeled with $[i+1, 2j]$ and $[i+1, 2j+1]$ (as defined in Fig. 7a). To do so, we choose a cost function I, such that if $I_{i,j} \geq I_{i+1,2j} + I_{i+1,2j+1}$, we favor block $[i, j]$, and vice versa. A cost function commonly used for this task, is the Shannon Information Function, which corresponds to minus the Entropy function (Coifman and Wickerhauser, 1992; Saito, 1994; Mallat, 1998):

$$I_{i,j} = \sum_{k=1}^{m} q_{i,j,k} \log(q_{i,j,k}) \tag{4}$$

where

$$q_{i,j,k} = \frac{|\langle f(t),\ \psi_{i,j,k}\rangle|^2}{\|f(t)\|^2} \tag{5}$$

is the energy contribution of $f(t)$ in packet $\psi_{i,j,k}$.

Once we have calculated I for all the blocks in the decomposition, we can perform an overall maximization of I, and the resulting set of blocks is the basis of maximum information, and thus defined as the best basis. The maximization is performed as follows: starting at the bottom level $i=L$ (the lowest row in Fig. 7a), we compare every pair of blocks with their corresponding block at level $i=L-1$ and choose according to the maximum value of I. After choosing the blocks from these two levels, we go up one level and compare the blocks of levels $i=L-1$ and $i=L-2$. We repeat this procedure up to the top level $i=1$.

There is an intuitive relation between the entropy cost function and the physical entropy. Maximum entropy corresponds to an homogeneous distribution over micro-states. Thus, minimum entropy (i.e. maximum

information) should correspond to concentration of energy over a minimum of basis elements.

4.2. Local discriminant basis

The LDB method (Saito, 1994) is a generalization of the best basis algorithm. It is used when the main point is to enhance the differences between two signals. In resemblance to the best basis, the LDB is a basis that uses the smallest number of significant basis elements to separate two signals.

The scheme of the LDB is as follows: given two signals, $f^{(1)}(t)$ and $f^{(2)}(t)$, we build for each signal the block diagram (as described before). Next, we choose the cost function M to be the Shannon's mutual information (minus the mutual entropy) (Saito, 1994):

$$M_{i,j} = \sum_{k=1}^{m} q^1_{i,j,k} \log \frac{q^1_{i,j,k}}{q^2_{i,j,k}} + \sum_{k=1}^{m} q^2_{i,j,k} \log \frac{q^2_{i,j,k}}{q^1_{i,j,k}} \quad (6)$$

where

$$q^1_{i,j,k} = \frac{|\langle f^{(1)}(t),\ \psi_{i,j,k}\rangle|^2}{\|f^{(1)}(t)\|^2} \qquad q^2_{i,j,k} = \frac{|\langle f^{(2)}(t),\ \psi_{i,j,k}\rangle|^2}{\|f^{(2)}(t)\|^2} \quad (7)$$

Next, we perform an overall maximization of the mutual information M in a way similar to the maximization of I in the best basis algorithm.

In some cases, to reduce the influence of fluctuations, M is calculated not on the energy distribution of a signal ($q_{i,j,k}$), but on an averaged measurement. We take N^1 measurements of signal type 1 $\{f^{(1)}_n\}_{n=1}^{N^1}$, and calculate the energy distribution over the packets:

$$\Gamma^1_{i,j,k} = \frac{\sum_{n=1}^{N_1} |\langle f^{(1)}_n,\ \psi_{i,j,k}\rangle|^2}{\sum_{n=1}^{N_1} |f^{(1)}_n|^2} \quad (8)$$

In this case, the discriminant measure between the two signal types, is calculated on $\Gamma^1_{i,j,k}$ and $\Gamma^2_{i,j,k}$, and we continue in the same manner.

5. Applying the method for neural spikes

We apply our method on measurements of in vitro cell cultures using multi electrodes arrays (Segev et al., 2001). The neurons are randomly attached to the surface. The capacitance coupling between the neurons' membrane and the electrode enables the recording of the spontaneous firing of neural spikes (Jimbo et al., 1993, 2000; Kamioka et al., 1996; Camepari et al., 1997; Maher et al., 1999). The expression of the neural spikes in the recording is affected by the distance from the center of the electrode and the quality of the coupling (Lewicki, 1998; Rinberg et al., 1998).

The recording of the neuronal activity is held consecutively for days, and we have cases of as many as five neurons read by a single electrode. The signals collected from the electrodes are digitized at a rate of 12 kHz. The WPD of an event is performed on a window of 128 samples ($\sim$10 ms) with the spike at it's center (as a consequence, $L = \log_2(N) = 7$).

The first step is to select an appropriate mother wavelet. The one we have selected is the 3rd order Coiflet. Fig. 8 presents different packets of this mother wavelet. These Coiflets are nearly symmetric and have little overlapping with their neighbors both on time and frequency domains. The compactness is important, as it does not spread the signal's energy into neighboring packets (see Mallat, 1998 for discussion).

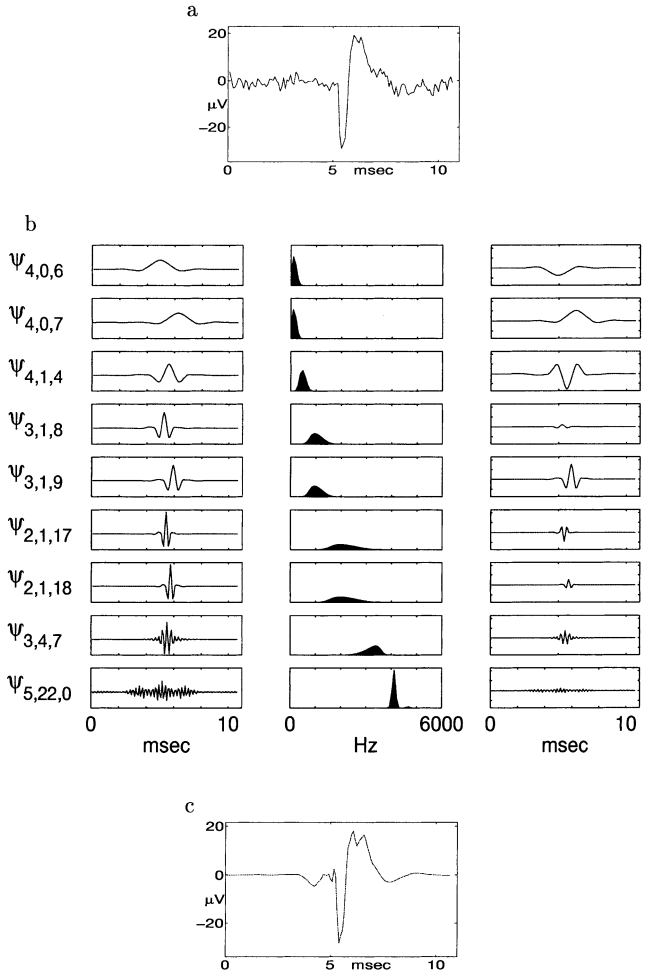

Fig. 8. WPD of spike using 3rd order Coiflets: (a) a typical spike. (b) The decomposition of the spike into nine selected packets. Each row presents the time and spectral behavior of a packet (left and middle, respectively), and the packet multiplied by the coefficient of decomposition (hence $\psi_{4,0,6}$ and $\psi_{4,0,7}$ carry opposite signs etc.). (c) The reconstruction of the spike from the nine packets' coefficients. It is clear that the packets grab the main shape of the spike and most of its energy. It is also clear that the usage of the nine packets filters most of the noise in which the neural spike is embedded.

5.1. Detecting spikes using wavelet packets

Whenever attempting to detect neural spikes merely with a threshold detector, there is a trade off between the level of the threshold and acceptance of noise events. If the threshold is set at a high level, very few noise events will be detected. However, low amplitude spikes will be a priori excluded. This can be mended by lowering the threshold to include the low amplitude spikes accompanied by noise events, and using a well performing sorting method to disqualify the noise. In this sense, a good sorting method can be adopted also to detect spikes whose amplitudes are comparable with the noise level.

Moreover, in cases of rapid firing of spikes (such as in Fig. 1c), a simple real-time threshold detector is not capable of detecting all the peak amplitudes due to the frequent firing. In order to analyze all the spikes that appear in the recording, and not just those that appear in the center of a digitized window, we have developed a method, using wavelet packets, to mark all the peak amplitudes in a window. The procedure is as follows.

We decompose the 20 ms digitized window using all the shifted versions of the wavelet packet $\psi_{2,0}$ (which by applying the best basis technique was found to be the most characteristic packet for detecting a spike, regardless of its origin). On this decomposed signal, we search for all the samples that are local minima, that is-extremum among its 12 nearest neighbors (0.5 ms to either side). Each of these peak amplitudes is suspected to be a spike, and a window of 128 samples ($\sim$10 ms) is isolated with the peak at its center for further WPD analysis and sorting. Fig. 9 presents an example of this procedure. This procedure is motivated by the shift-invariant WPD (for further mathematical description see Cohen et al., 1997; Zouridakis and Tam, 1997). Note that by doing so, we confine our method to detect spikes that are at least 0.5 ms apart.

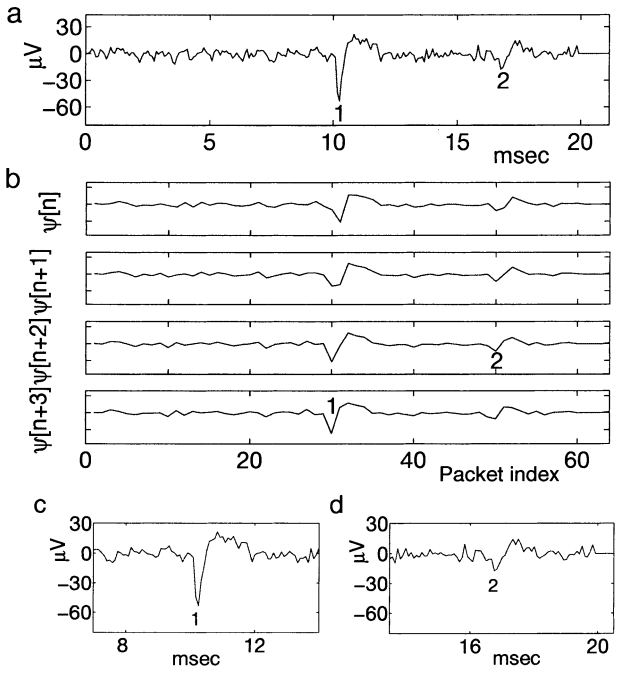

Fig. 9. Isolating the exact location of amplitude peaks (suspected spikes) in an event. (a) A recorded event with two spikes (256 samples in length). (b) Decomposition of the event using the wavelet packets $\psi_{2,0,k}[n]$, $\psi_{2,0,k}[n+1]$, $\psi_{2,0,k}[n+2]$ and $\psi_{2,0,k}[n+3]$. The packet $\psi_{2,0,k}$ is four samples in width, hence $0 \le k < 64$ for each of the four shifted version of the packet. (c) The peak at 10.25 ms (marked by 1) is best covered by $\psi_{2,0,30}[n+3]$ (since $\psi_{2,0,30}[n+3]$ is both an extremum compared with $\psi_{2,0,30}[n]$, $\psi_{2,0,30}[n+1]$, $\psi_{2,0,30}[n+2]$ and to its 12 nearest neighbors $\psi_{2,0,k}[n+3]$ $k = 24, \ldots, 36$). The packet $\psi_{2,0,30}[n+3]$ is related to sample 123 in the original event (30 translations of length 4 added to an initial shift of three indices), hence the window covering the peak is centered around sample 123. (d) The peak at 16.75 ms (marked by 2) is best covered by $\psi_{2,0,50}[n+2]$, hence the window covering this peak is centered around sample 202.

5.2. Selecting wavelet packets for detection and sorting

To demonstrate the efficiency of our method, we worked on a recording from an electrode that measured neural spikes originating from five different neurons (as presented in Fig. 1c). We manually selected 100 events for each of the neurons and of noise. Fig. 10 presents the different spikes' groups.

We performed the WPD on a window of 128 samples around each event. Since LDB is designed to find the separating basis between two types of signals, we generalized the LDB technique in order to differentiate between four signal types. We apply the LDB procedure on each pair of groups and retrieved a basis separating best among the pair (ten bases for all the different couples). The union of these bases is an over-redundant discriminant basis between all the groups. To extract one non-redundant basis, we first checked which packets appeared in all the bases. We found that the union of those bases was only slightly over-complete, as most of the bases contained similar packets. For the rest of the packets, we selected the level of decomposition that was most common among the ten bases. The resulting basis is a discriminating basis for all six groups.

Among the 128 basis elements, the nine most discriminating packets were selected both for detection and sorting. Those packets are $\psi_{4,0,6}$, $\psi_{4,0,7}$, $\psi_{4,1,4}$, $\psi_{3,1,8}$, $\psi_{3,1,9}$, $\psi_{2,1,17}$, $\psi_{2,1,18}$, $\psi_{3,4,7}$ and $\psi_{5,22,0}$. Fig. 8 presents the temporal shape and frequency coverage of the selected packets.

In Fig. 8, we present the decomposition of a typical spike. The spike is decomposed into the nine packets, and then reconstructed. Note that the reconstruction of the spike using the nine packets reveals the unique features of the spike's shape and covers most of the spike's energy ($\sim$93% of the window's energy). It is possible to view the union of the selected packets also as

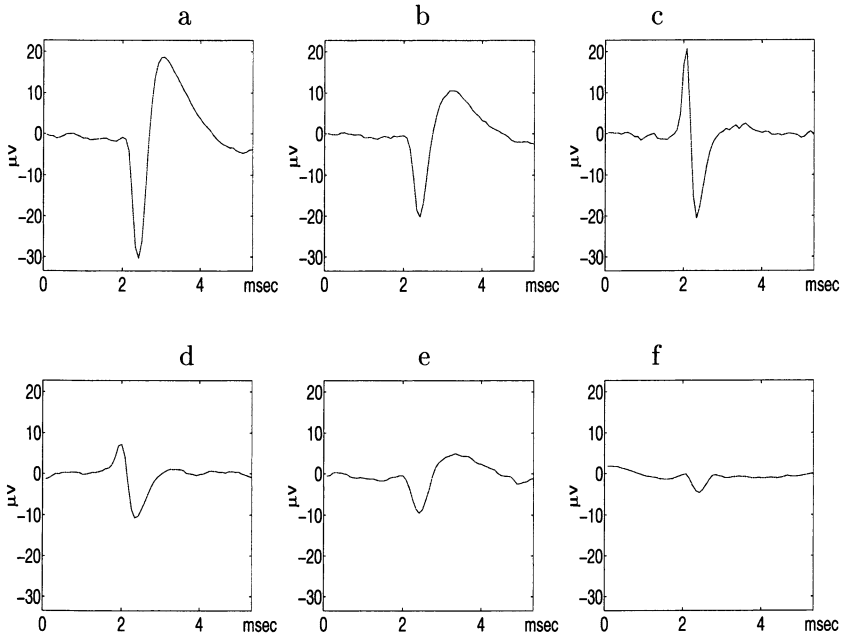

Fig. 10. Spike detection and sorting. The six sub-plots present the groups of the spikes in the data, and as sorted by the classification process. Note five different neural spike shapes (a–e) and the class of background noise (f).

a highly optimized filter for neural spikes, that preserves its shape and filters out most of the background noise.

Figs. 11 and 12 presents the process of detecting multiple spikes in the same window, where the spikes are from different neurons and also overlap. Each of the detected peaks is independently decomposed using the same nine packets, and is then reconstructed. Note that

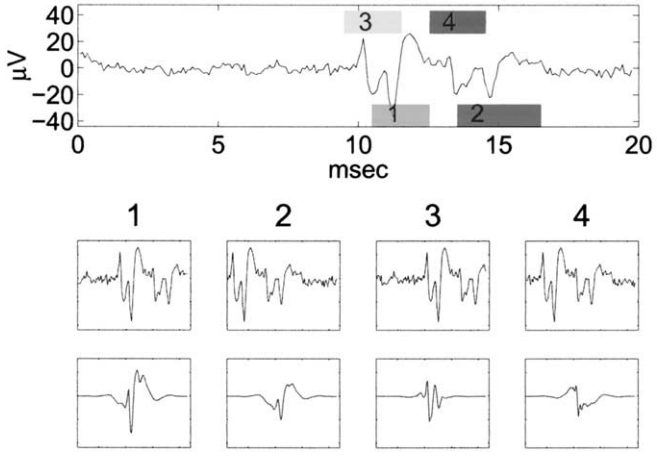

Fig. 11. WPD in a window with a number of spikes. The top level presents an event with a number of spike events. In the second level, we present a copy of the 128 samples with an event at the center. In the third level, we present the reconstruction of the spike using the nine packets. Note that we succeeded in detecting two different but overlapping spikes (marked by 1 and 3), and sorted them both from the spike marked by 2 and 4. The later two spikes are also very close to one another, hence the shape of spike 4 is a-priori not easy to classify. However, even though the overlapping of the spikes damages the typical shape of the spike, the packets capture the key features of all the spikes in the event.

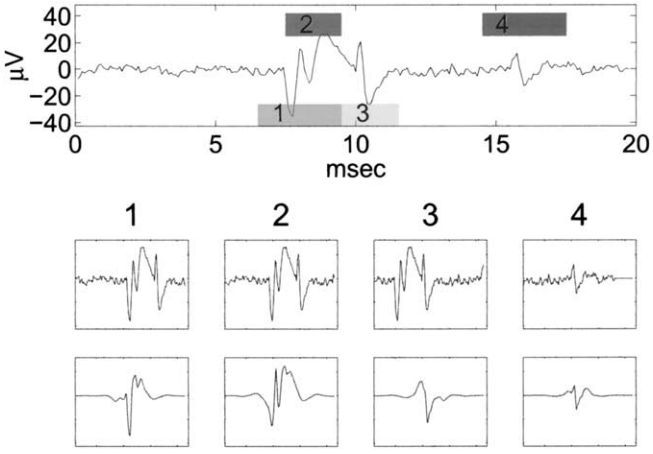

Fig. 12. Another example of a window with a number of spikes. Note that even though spikes 1 and 2 are overlapping, we succeeded in detecting both spikes. Moreover, the shape of the reconstructed spikes is almost not affected by the overlapping.

the packets succeed in preserving three different spike shapes in this window even though two of them overlap.

In order to visualize the separation ability of the selected packets, we embed the coefficients of those packets in a 9 dimensional space. Separation means that spikes from each type are clustered together in distance from clusters of other spike types. In Fig. 13 we present a matrix plot of four of the packets' coefficients. Projections of the clusters on the different planes are plotted on each sub-figure. It is clear that the group of noise events is completely isolated from the five spike

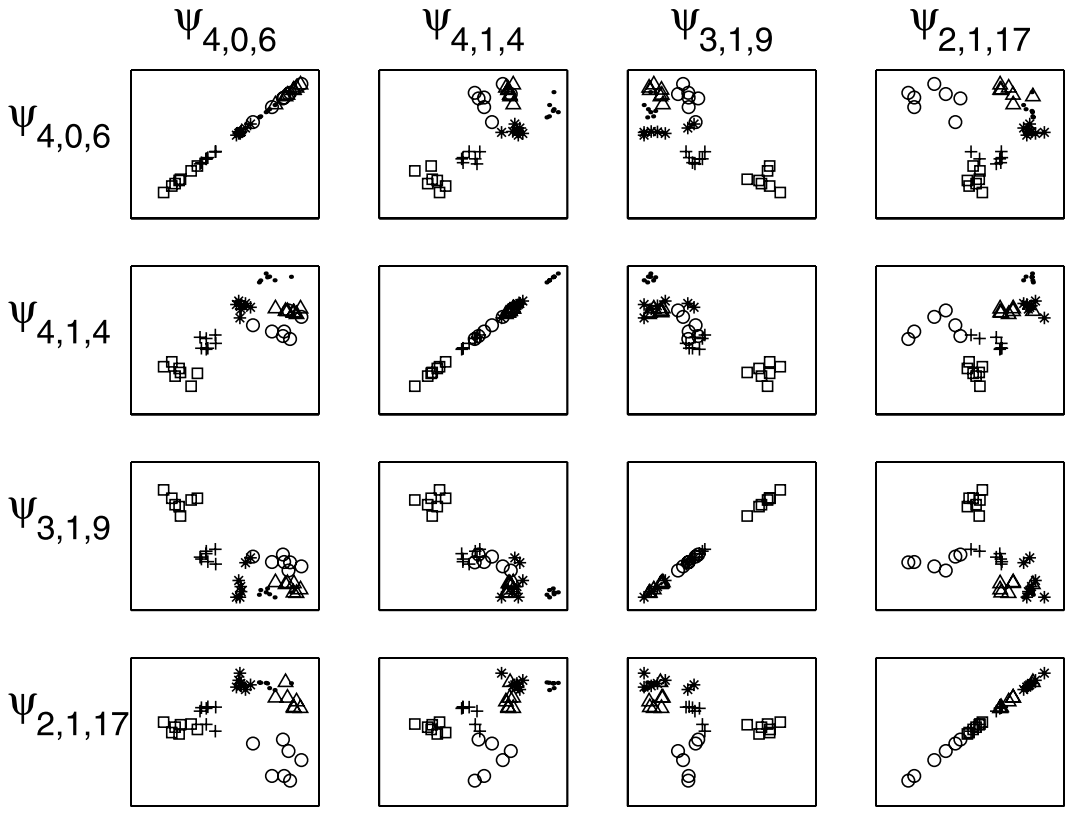

Fig. 13. Matrix plot of wavelet packet clusters of different spike type. Each sub-plot shows the projection of the coefficient space onto pairs of coordinates. The k-means classifier was used to cluster the wavelet packets coefficients. There is a good discrimination among the different clusters: small dots represent noise and the five shapes are different spike types. The units are arbitrary.

types, while the different spike types are also well separable.

5.3. Sorting into neural classes

As the different spike types and noise are clustered in the 9 dimensional space, we can automatically separate the clusters by using a classifier such as the k-means classifier (Bow, 1991; Dellaert, 1999). The k-means algorithm finds the N_k most distinct clusters in a multi-dimensional space (where N_k is given as an input to the algorithm). We use the Euclidean measure to calculate the distances between vectors in the 9 dimensional space. The k-means is not designed to find the optimal number of clusters in the data. In our case, N_k is influenced by the number of spike types in the recording and thus may vary according to the number of neurons measured by the electrode. As a consequence, N_k is determined manually for each tested electrode at the beginning of the process. This is done by applying the k-means classifier on a short segment of the recording (first 2000 events) with different selections on N_k. The minimal N_k that produces satisfying results (all the spike types are separated from the noise and among each other), is selected for that electrode. In many cases, satisfying results were achieved for N_k larger than the number of spike types + one for noise. That is because the noise has different forms that are some times more pronounce than small amplitude spikes.

These 2000 events comprise the training set for the automatic classification of events throughout the recording. Each event is classified according to its Euclidean distance from the k centers of clusters in the training set. The cluster that is closest determines the class of the event.

In Fig. 14a we present the centers of the five clusters of spikes found in the training set using k-means. Note the resemblance to the mean spikes presented in Fig. 10. Table 1 presents the detection and sorting results on a test set of 400 spikes and 400 noise events. One hundred of the spikes are in overlap with a nearby spike (distance less than 1.5 ms). An overall of 92% of the spikes in the test set were detected and 87% were classified correctly. There were less than 2% false detections of noise. Regarding the overlapping spikes, 74 of the 100 were detected, 66 of them were classified correctly. The verification of the classification results was performed manually.

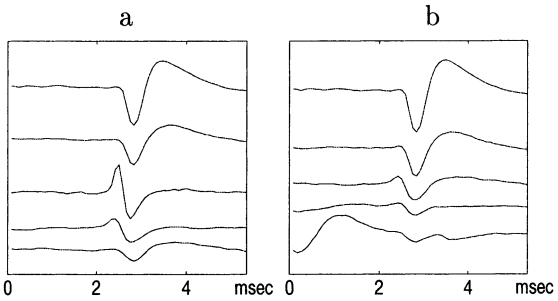

Fig. 14. Comparison of the clustering performance of the WPD method and the principal components. The centers of the five best isolated clusters of our nine packets (a) and the top two principal components (b). The clustering is done using k-means for both methods. Our method can sort the five different spike types, with high resemblance to Fig. 10. The principal components present four possible spike clusters (top four curves in b), but the fifth cluster stands for a typical lag between consecutive spikes. Hence, principal components fail to sort among the different spikes types.

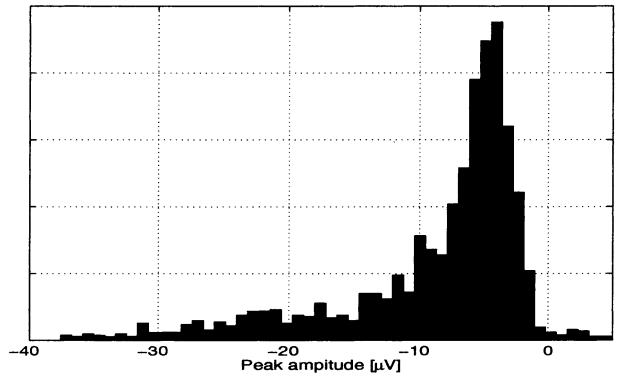

Fig. 15. A histogram of the peak amplitude of events. Those are the same events that appear in Figs. 10 and 14. It is clear that there is no clear separation of the different spike types from the noise group.

6. Comparison with other methods

Various signal processing methods have been used to extract the features of neural spikes in the process of detection and sorting (see Lewicki, 1998; Wheeler, 1999 for reviews). One family of methods, is to describe each spike by a set of features. Simple features that are known to be used for describing spikes are the maximal and minimal peaks, the width of the spike etc. Fig. 15 present a histogram of the peak amplitude of events. Those are the same events that comprise the matrix plot in Fig. 13. It is clear that there is no clear separation of the five spike types from the noise group.

Another family of methods, is to filter the data utilizing a series of templates of spike waveforms (Lewicki, 1998). Variants of this method based on the principal components analysis (PCA) have been developed (Lewicki, 1998). The concept of PCA is to transform the data into a new set of coordinates that signify the direction of largest variations in the data. These directions (named principal components) are known to be the eigenvectors of the covariance matrix of the data (Glaser, 1971). The eigenvectors form an orthogonal basis, and the magnitude of the relative eigenvalues designates the importance of the components in representing the data. The top principal components are expected to enhance the differences among the shapes of different spike types.

There are a vast variety of methods used by researchers for feature extraction of neural spikes. We compare our sorting results of our method with the widely used PCA (presented by Lewicki, 1998), and to a method somewhat related to our own, recently presented by Zouridakis and Tam (1997) that uses the WT coefficients rather than WPD coefficients. As will be shown further in this section, our method is superior to these methods both in detecting and in sorting overlapping spikes. Moreover, as stated in the introduction, in this manuscript we do not compare different clustering methods, but rather demonstrate the efficiency of our method merely using the simple k-means classifier. The effect of different clustering methods on the spike

Table 1
Performance of detection and sorting of the WPD method compared with WT and PCA

	WPD						WT						PCA					
	C_1	C_2	C_3	C_4	C_5	X	C_1	C_2	C_3	C_4	C_5	X	C_1	C_2	C_3	C_4	C_5	X
N_1	54	1					46	7	2				51	4				
N_2	3	89		1	2		1	79	6		8	1	6	80	5			4
N_3		1	30	4		5			27	8		5		5	15	3	2	15
N_4				92	3	19			5	84	9	16			40	21	3	50
N_5		2		3	81	8		3		14	64	13		5	69	3	3	14
FA				4	3					19	8				1	75	59	

Each row presents the classification of neural spikes (named N_i where $i = 1, \ldots, 5$), and the last row (FA) presents noise events that were falsely detected and classified as neural spikes. For each sorting method, five columns present the classification distribution among the five classes, and the last column (X) states the number of spikes that were not detected (classified as noise). The classification was performed using k-means for all methods. For WT we used the nine most energetic wavelets and for PC we used the top two principal components. The results in this table are verified manually.

sorting efficiency will be presented elsewhere. We intend to present our work on different classifiers emphasizing the clustering problem elsewhere.

The comparison is performed using the same data that was described in Section 5. In Fig. 16a we show the comparison between the separability of the principal components and the wavelet packets. Separation ability is defined via the capability to capture a higher portion of a spike's energy than of noise energy. Thus, the 'bending' point in the energy distribution curve distinguishes between the separating and non-separating packets (or principal components). We find that only the first two principal components have separation ability. On the other hand, there are nine separating wavelet packets. The first two principal components contain on average 86% of the strongest spikes' energy, while the nine packets capture an average of 84%. However, the principal components contain more noise energy than the packets (Fig. 16b). Using the WT, nine wavelets seem to have separation ability, and they capture 81% of the energy. As will be shown ahead, the fact that only two principal components have separation ability in a data set of five different spikes, will ultimately prevent the PCA method from detecting and sorting all the spike types.

As for the WPD case, the separation of clusters is done using the k-means classifier, and the sorting of a test set is done as described previously. In Fig. 14b we present the centers of the five clusters found in the test set using the principal components. It is clear that the principal components detect three spike templates, but fail to distinguish the two weaker classes from the noise. Using the WT, the centers of the clusters follow those of the WPD (Fig. 14a).

Table 1 presents the comparison of detection and sorting results on a test set of 400 spikes and 400 noise events. One hundred of the spikes are in overlap with a nearby spike (distance less than 1.5 ms). For WT using nine wavelets, an overall of 91% of the spikes in the test set were detected and 75% were classified correctly. There were 7% false detections of noise. Regarding the overlapping spikes, 73 of the 100 were detected, 45 of them were classified correctly. For PCA using the two separating components, an overall of 59% of the spikes in the test set were detected and 42% were classified correctly. There were 34% false detections of noise. Regarding the overlapping spikes, 50 of the 100 were detected, 26 of them were classified correctly. The verification of the classification results was performed according to the a-priori knowledge of the spikes origin. That is, the tested data set was carefully inspected by an independent trained eye in order to determine the ascription of the spikes to the different groups. The superiority of our method compared with the PCA is outstanding in all aspects (detection and sorting of both overlapping and non-overlapping of spikes and in rejection of noise). Compared with the WT, we notice substantial improvement in noise rejection, and in the correct sorting among the five neurons (from 75 to 87%). The contribution is, as expected, especially in the case of overlapping, where the sorting performance improved from 45 to 66%.

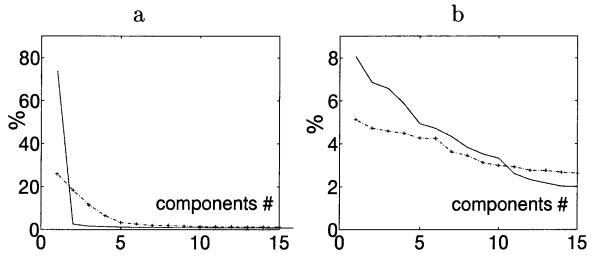

Fig. 16. A comparison between the separation ability of the wavelet packets (dashed lines) and principal components (solid lines). (a) Distribution of energy percentage over the coefficients. Only the first two principal components have separation ability as opposed to nine wavelet packets. (b) Distribution of noise energy: the first principal components capture more noise energy than the wavelet packets, thus reducing the ability to separate spikes from noise. The trade off between covering the signal's energy and excluding noise energy determines the number of separating components.

7. Summary and conclusions

We have presented a new method for detection and sorting of neural spikes using WPD. We explained how to select wavelet packets with separation ability among a large number of spikes using the LDB algorithm. We have stated that the performance of this feature extraction method enables us to use a simple clustering method such as the k-means algorithm. We demonstrated the efficiency of our method on in vitro neural networks recordings, and compared the performance to those of the principal components method and regular WT. The main advantage of our method is the ability to detect and sort overlapping spikes. Furthermore, our method is capable of sorting a large number of neurons on a single electrode. The method is general and can be adapted for recordings of in vitro tissues as well as in vivo preparations. The use of more sophisticated classifiers (e.g. support vector machines), will enhance the sorting performance.

During long term recordings, the neurons move with respect to the electrodes. As a result, the shapes of the recorded spikes variate. Our method is capable of adaptively following these variations, and we shall present this concept elsewhere.This feature extraction method also enables us to investigate the inner structure of the neural action potential, as it enables to reliably describe the spike as a linear combination of nine smooth wavelet packets.

Acknowledgements

We have greatly benefited from conversations with A. Averbuch, M. Benveniste, and Y. Shapira. This research has been partially supported by a grant from the Israeli Academy of Sciences, the Sackler Institute, the Adams Super Center for Brain Studies, and the Kodesh Institute. One of us (Ronen Segev) is a recipient of the Intel Ph.D. fellowship.

References

Bow S. Pattern Recognition and Image Processing. Marcel Dekker, 1991.

Camepari M, Bove M, Maede E, Cappello M, Kawana A. Experimental analysis of neuronal dynamics in cultured cortical networks and transitions between different patterns of activity. Biol Cyber 1997;77:153–62.

Cohen I, Ras S, Malah D. Orthonormal shift-invariant wavelet-packet decomposition and representation. Signal Proc 1997;57:251–70.

Coifman R, Wickerhauser M. Entropy-based algorithms for best basis selection. IEEE Trans Inform Theory 1992;38(2):713–8.

Coifman R, Meyer Y, Wickerhauser M. Wavelet analysis and signal processing. In: Wavelets and Their Applications. Boston, 1992.

Daubechies I. Ten Lectures on Wavelets. SIAM, 1992.

Dellaert F. *K*-means MATLAB function, http://www.cs.cmu.edu/people/dellaert/, 1999.

Donoho D, Huo X, Yu T. MATLAB toolbox, http://www-stat.stanford.edu/ wavelab/, 2000.

Egert U, Schlosshauer B, Fennrich S, Nisch W, Fejtl M, Knott T, Muller T, Hammerle H. A novel organotypic long-term culture of the rat hippocampus on substrate-integrated multielectrode arrays. Brain Res Proto 1998;2:229–42.

Fromherz P, Muller B, Stett A. Two-way silicon-neuron interface by electrical induction. Phys Rev E 1997;55:1779.

Glaser E. Separation of neuronal activity by waveform analysis. Adv Biomed Eng 1971;1:77–136.

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

Jimbo Y, Robinson H, Kawana A. Simultaneous measurement of intracellular calcium and electrical activity from patterned neural networks in culture. IEEE Trans Biomed Eng 1993;40:804–10.

Jimbo Y, Kawana A, Parodi P, Torre V. The dynamics of neuronal culture of dissociated cortical neurons of neonatal rats. Biol Cyber 2000;83:1–20.

Kamioka H, Maeda E, Jimbo Y, Robinson H, Kawana A. Spontaneous periodic synchronized bursting during formation of mature patterns of connections in cortical culture. Neurosci Lett 1996;206:109–12.

Letelier J, Wever P. Spike sorting based on discrete wavelet transform. J Neurosci Methods 2000;101:93–106.

Lewicki M. Bayesian modeling and classification of neural signals. Neu Com 1994;6:1005–30.

Lewicki M. A review of methods for spike sorting: the detection and classification of neural action potentials. Net Com Neu Sys 1998;9:R53–78.

Maher M, Pine J, Wright J, Tai Y. The neurochip: a new multi-electrode device for stimulating and recording from cultured neurons. J Neurosci Methods 1999;87(1):54.

Mallat S. A wavelet tour of signal processing. Academic Press, 1998.

Rinberg D, Davidowitz H, Tishby N. Multi-electrode spike sorting by clustering transfer functions. Proceedings of the Neural Information Processing Systems, 1998.

Saito N. Local Feature and Its Application Using a Library of Bases. Ph.D. Thesis, Yale University, New Haven, 1994.

Segev R, Shapira Y, Benveniste M, Ben-Jacob E. Observations and modeling of synchronized bursting in 2d neural networks. Phys. Rev. E 2001;64:011920.

Stenger D, Mckenna T. Enabling Technologies for Cultured Neural Networks. Academic Press, 1994.

Unser M, Aldroubi A. A review of wavelets in biomedical applications. IEEE Proc 1996;84:626.

Wheeler B. Automatic Discrimination of Single Units. CRC, 1999.

Yang X, Shamma S. A totally automated system for the detection and classification of neural spikes. IEEE Trans Biomed Eng 1998;35:806.

Zouridakis G, Tam D. Multi-unit spike discrimination using wavelet transforms. Comput Biol Med 1997;27:9–18.

Zouridakis G, Tam D. Identification of reliable spike templates in multi-unit extracellular recordings using fuzzy clustering. Comput Methods Prog Biomed 2000;61:91–8.