J Comput Neurosci (2015) 38:249–261
DOI 10.1007/s10827-014-0539-z

Spike detection methods for polytrodes and high density microelectrode arrays

Nicholas V. Swindale · Martin A. Spacek

Received: 25 June 2014 / Revised: 24 September 2014 / Accepted: 4 November 2014 / Published online: 20 November 2014

Abstract This paper compares the ability of different methods to detect and resolve spikes recorded extracellularly with polytrode and high-density microelectrode arrays (MEAs). Detecting spikes on such arrays is more complex than with single electrodes or tetrodes since a single spike from a neuron may cause threshold crossings on several adjacent channels, giving rise to multiple events. These initial events have to be recognized as belonging to a single spike. Combining them is, in essence, a clustering problem. A conflicting need is to be able to resolve spike waveforms that occur close together in space and time. We first evaluated three different detection methods, using simulated data in which spike shape waveforms obtained from real recordings were added to noise with an amplitude and temporal structure similar to that found in real recordings. Performance was assessed by calculating the percentage of correctly identified spikes vs. the false positive rate. Using the best of these detection methods, two different methods for avoiding multiple detections per spike were tested: one based on windowing and the other based on clustering. Using parameters that avoided spatial and temporal duplication, the spatiotemporal resolution of the two methods was next evaluated. The method based on clustering gave slightly better results. Both methods could resolve spikes occurring 1 ms or more apart, regardless of their spatial separation. There was no restriction on the temporal resolution of spike pairs for units more than 200 μm apart.

Action Editor: Jonathan David Victor

N. V. Swindale (✉) · M. A. Spacek
Department of Ophthalmology and Visual Sciences,
University of British Columbia, 2550 Willow St., Vancouver, B.C.,
Canada V5Z 3N9
e-mail: swindale@mail.ubc.ca

Keywords Spike sorting · Event detection · Polytrodes · Multi-electrode arrays · Clustering · Tetrodes · Neurophysiological recording

1 Introduction

Extracellular electrodes that have many (8 or more) closely spaced individual recording sites are capable of recording spike trains from large numbers of neurons with high temporal and spatial accuracy (Bragin et al. 2000; Buzsáki 2004; Blanche et al. 2005). Such recordings offer an unprecedented ability to relate the spiking activity of populations of single neurons to sensory stimuli and behaviour. An essential stage, preliminary to identifying the signals from individual neurons in such recordings, is the initial detection of candidate electrical events that are likely to originate from neuronal action potentials (Chandra and Optican 1997; Lewicki 1998; Rebrik et al. 1999; Obeid and Wolf 2004; Choi et al. 2006; Maccione et al. 2009). Almost all methods for doing this require the application of a threshold to some measure derived from a voltage waveform at a specific point in time. A variety of problems arise in applying such thresholding procedures to recordings made with MEAs. These electrodes, by design, have recording sites spaced sufficiently closely that a single neuronal spike will often give rise to threshold crossings detectable on several adjacent channels. These signals need to be recognized as belonging to the same event (Jäckel et al. 2012; Swindale and Spacek 2014) which is essentially a clustering problem. If this does not happen correctly, duplicate events will be produced (more can occur, though duplication is the most common error in practice) which may give rise to errors in subsequent processing. Spatiotemporal windowing can be used to exclude duplicates, however the use of too large a window will reduce the ability to resolve spikes that occur close together in space and time. A final problem, common to

all extracellular recording approaches, is to be able to maximize the percentage of low amplitude spikes detected while minimizing the number of false positives. This procedure in principle can be aided by prior knowledge of spike waveform shape, however the biophysical and experimental factors that give rise to extracellularly recorded waveform shapes are complicated and only partly understood (Gold et al. 2006). In practice, extracellularly recorded spikes can have varying numbers (1–4) of peaks of alternating sign and varying relative magnitude. The shapes may also be impacted by filtering artefacts (Wiltschko et al. 2008; Quian Quiroga 2009) which may add, or accentuate, subsidiary small magnitude peaks. Hence it would seem desirable not to make strong assumptions about spike shape in designing detection methods beyond the idea that one or more peaks of varying sign and magnitude may be present.

These problems were addressed as follows. We generated simulated recordings by taking averaged multi-channel spike waveforms (templates) obtained after spikes in real recordings had been sorted and adding them to artificial background noise. The noise had an autocorrelation structure similar to that of real spike-free recording noise. Using these simulations, we then compared the ability of a variety of event detection methods to detect small amplitude spikes. Taking the best of these methods, we then compared two methods for avoiding event duplication, one based on spatiotemporal lockout (STL) and the other based on spatiotemporal clustering of events, a method here termed proto-event clustering (PEC). We estimated spatial and temporal window parameters that seemed likely to avoid event duplication of the majority of types of spike waveform. Taking these parameters, we then estimated the spatiotemporal resolution of the STL and PEC methods. Finally, using real recordings, we compared the processing time of both methods.

2 Methods

2.1 Physiological data

Spike shapes and noise samples were obtained from data recorded with 54 channel polytrodes (University of Michigan Center for Neural Communication Technology and NeuroNexus) inserted into the visual cortex of anesthetized cats. Electrodes were either of 2 or 3 column design with sites 50, 65 or 75 μm apart. Experimental procedures are described more fully in Blanche et al. (2005) and Swindale and Spacek (2014). Experiments were carried out in accordance with Canadian Council on Animal Care guidelines and institutional protocols approved by the Animal Care Committee of the University of British Columbia. Single periods of continuous data acquisition lasted typically from 15 to 45 min and consisted of recordings of spontaneous as well as visually driven activity. The signals coming from the recording channels on the electrode were amplified and band pass filtered from 500 Hz–6 KHz before being digitized with 12 bit resolution at a rate of 25 KHz (a sampling interval of 40 μs).

2.2 Simulated recordings

Noise with a temporal autocorrelation function that was similar to that of real background noise in recordings was generated as

$$V(n,t) = \xi(n,t) + \sum_{\tau=1}^{30} V(n,t-\tau)W(\tau) \quad (1)$$

where V is the signal voltage, n is channel number, t and τ are time sample points and ξ is a random number with a Gaussian distribution. The coefficients $W(\tau)$ were derived from the autocorrelation function of periods of real data lacking detected events. The variance of ξ was chosen to give a variance of V equal to approximately 10 μV, similar to values measured in the real recordings. Figure 1 shows autocorrelation functions calculated from spike-free periods in 3 different animals, together with the autocorrelation function of the simulated noise.

For tests of detection, two sets of 6 multichannel spike waveforms were chosen (Fig. 2) encompassing a variety of shapes found in our recordings. Spikes were sorted based on clustering of the principal component values derived from the aligned spike waveforms using only the channels on which the average spike waveform exceeded a defined threshold (Swindale and Spacek 2014). Each waveform (referred to here as a 'template') was the average of all the spikes from the sorted unit in the recording (this number was at least 120 and frequently considerably more) taken over all 54 electrode channels. Templates were calculated for a temporal window extending from −0.4 to 1.6 ms relative to the center of the template. The center was defined as the position of the initial, negative trough on the channel on which the peak-to-peak voltage was a maximum (subsequently referred to here as the center channel of the template). This position is indicated by the horizontal positions of the vertical axes in Fig. 2. Center channels are indicated by asterisks. The first set (S1 – S6) varied in terms of the relative size, number, and order of negative and positive peaks, which are factors that are likely to have an impact on detection. They also varied in their spatial width and temporal duration. This set was scaled to specific peak-to-peak heights on the center channel. It was used for tests of detection accuracy as a function of spike amplitude. The second set (S7 – S12) was used for tests of event duplication and spatiotemporal resolution. The templates were selected to include waveforms that were spatially and/or temporally extensive and which seemed particularly

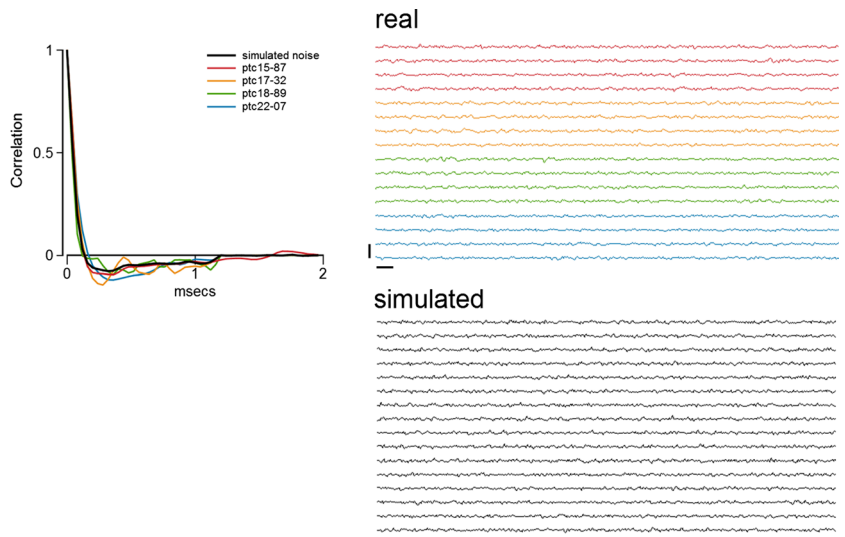

Fig. 1 *Left panel*: autocorrelation functions of recorded voltages from spike-free periods of recording from 4 different animals. The average of these functions was used to compute the simulated noise for tests of event detection. The *solid black line* shows the autocorrelation function derived from a period of simulated noise. *Right panels, upper*: periods of spike-free noise from the 4 recordings colored according to the scheme used in the *left panel*. Groups of four traces come from a series of positions along the polytrode, from the *top, upper middle, lower middle* and *bottom*. *Lower right panel* shows the simulated noise. The real noise had a median noise value (Eq. 2) of 9 μV. The simulated noise in this figure, and in all subsequent tests with it, had a median noise of 9.7 μV. Sixteen channels (out of 54) are shown. *Scale bar* shows 1 ms and 200 μV

likely to give rise to duplicate events and to prove challenging for tests of spatiotemporal resolution. Tests done on this second set did not require rescaling of the voltage values.

For measurements of detection accuracy vs. false-positive rate, height-accuracy curves and event duplication tests, we used periods of 1 or 5 min of simulated noise to which 1000 test waveforms were added at fixed intervals (to avoid spikes overlapping by accident). The times and center channels of each added spike were used to decide whether a subsequently detected event was correct or a false positive. For tests of spatiotemporal resolution, a pair of multichannel spike templates was chosen; 10,000 waveforms from each spike (i.e. 20,000 in all) were then added, at a given fixed spatial offset, to a 5 min period of simulated noise. The first spike of the pair was added at fixed, regular intervals (of 30 ms) and the second spike was added with a random temporal offset relative to the first spike, in the range −3.0 to +3.0 ms. After running event detection, the detected events were matched with the original ones to create two spike trains. The cross-correlation function of the two trains was then calculated using 0.2 ms bins and expressed as the percentage of the actual number of spike pairs present in a given bin. On average, this number was 333 (10,000 divided into 30 (0.2 ms) bins).

2.3 Event detection

We compared three different detection methods:

1. Bipolar threshold (BP): an event was registered at time t on a channel n if $|V(n,t)| \geq T$ where T is the threshold voltage. Following Quian Quiroga et al. (2004) we defined the threshold separately for each channel as a multiple, θ, of the channel noise, measured as the median of the absolute values of V, divided by 0.6745, i.e.

$$T = \frac{\theta \times median(|V|)}{0.6745}. \quad (2)$$

2. Non-linear energy operator (NEO): following Kaiser (1990), Maragos et al. (1993) and Kim and Kim (2000) an event was registered at time t on channel n if

$$V(n,t)^2 - V(n,t-1)V(n,t+1) \geq T^2. \quad (3)$$

This is also known as the Teager energy operator (Choi et al. 2006).

3. Following Blanche (2005) we implemented a dynamic multiphasic (DMP) detector. This requires first that a $V(n,t)$ be a local maximum, i.e. that $V(n,t-1) < V(n,t) \geq V(n,t+1)$ and that $V(n,t) \geq T$, or, conversely, that it is a local minimum and that $V(n,t) \leq -T$, and secondly that the absolute value of the difference between this voltage and at least one other voltage occurring within a temporal window $t - \delta$ to $t+\delta$ should equal or exceed $2T$. Figure 3 diagrams the procedure.

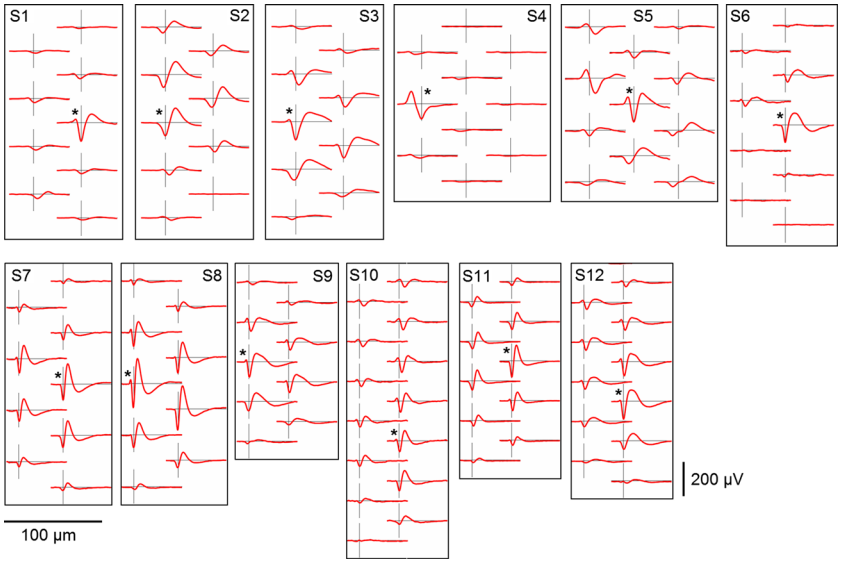

Fig. 2 The two sets of six spike waveforms used for the detection tests shown in Figs. 4 and 5 (S1 – S6) and for the tests of duplication and resolution shown in Figs. 6 and 7 (S7 – S12). Each waveform is the average of 120 or more spikes, taken from sorts of recordings from cat visual cortex as described in Swindale and Spacek (2014). For set S1 – S6 each spike is scaled to a peak-to-peak height of 80 μV; the vertical axis is 100 μV in all the panels and the horizontal axis is 1 ms for spikes S1 – S5 and 2 ms for *S6*. The waveforms are laid out in positions corresponding to the electrode site locations. For set S7 – S12 the heights are unscaled; the vertical axes are 200 μV and the horizontal axes are 2 ms. The intersection points of the axes are laid out according to the electrode site positions with the scaling indicated by the *scale bar*. The *vertical axes* also indicate the center point of each template, defined as the position of the negative trough of the waveform on the channel, indicated by an *asterisk*, for which the peak-to-peak amplitude is largest

2.4 Avoiding multiple events

The detection procedures described above will often yield multiple events per spike, distributed across several neighbouring channels. For example the DMP operator applied to the spike shown in Fig. 3 results in two events on the same channel. Some method is therefore needed for reducing these multiple events to just one per spike. We tested two methods for doing this, one based on spatiotemporal lockout and the other based on clustering.

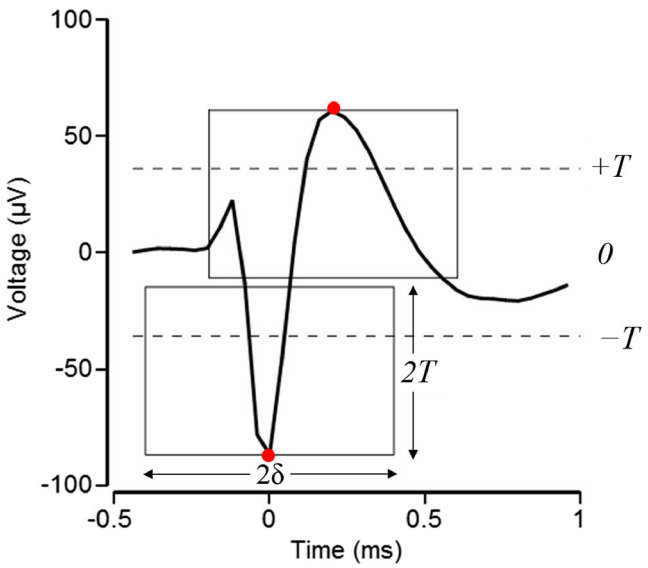

Fig. 3 Explanation of the dynamic multi-phasic event detection (DMP) method. Detection proceeds in two stages: first, points where a local voltage maximum (or minimum) lies above (or below) a threshold voltage, T, are identified (*red dots*); second, the voltage must change by an amount $\geq 2T$ within $\pm\delta$ ms of the initial detection time. This condition is satisfied if the waveform exits the bottom (or top) edge of a rectangle as shown in the Figure. The waveform illustrated gives rise to two detected events

2.4.1 Spatiotemporal lockout (STL)

This procedure vetoes events that occur within a defined spatiotemporal region around other events. This region can be defined by lockout parameters L_x and L_t. We typically used values of L_x in the range 100–200 μm and values of L_t in the range 0.5–1.0 ms. (The effects of particular choices of these values are examined below in Section 3.3). Two events were considered to overlap by this criterion if the corresponding event channels were less than a distance L_x apart and the event times were less than L_t apart. The lockout region is thus a rectangle with sides $2L_x$ and $2L_t$ centered on the event. One way of applying the method is to take a candidate event and register it as an event provided no previously registered event overlaps with it. However this may be inefficient because spikes may not necessarily be first detected on the channel on which the waveform amplitude is a maximum. It is desirable instead to center the lockout region, as far as possible, on the spatiotemporal center of the spike. For this reason, when an event was detected, a local search was made to find the

channel which had the largest peak-to-peak voltage, within the region defined by L_x and L_t. This was done by first finding a position midway between the maximum and minimum voltages found within$\pm L_t$ of the initial event time, and then finding the channel, within the set of channels$<L_x$ distant from the initial channel, on which the peak-to-peak voltage was a maximum. The resulting time and channel were then taken as the center position of the event. Then, if any previously registered events were found to be within the lockout region, the new event was discarded.

The array containing the voltages is two-dimensional, with one of the dimensions being channel number and the other being time. It is advantageous to search the array for events with channel number varying most rapidly as this results in events being detected in an approximately temporal order. The search of previously detected events for potential overlap with the new one can then be sped up by confining the search to a fixed number (we chose 200) of immediately preceding events.

2.4.2 Proto-event clustering (PEC)

A second way to avoid duplicate event detection is to store the times and channel numbers of *all* the detected events and then to cluster (or merge) them so that each cluster corresponds to a single spike. For clarity we will refer to these initially detected events as proto-events. With the DMP method, which confines detected events to local voltage maxima and minima, we usually found 1–10 proto-events per spike. Proto-events were stored as a list of points, $\mathbf{p}_j$ (j =1 … J) with scaled space-time coordinates (x_j/σ_x, y_j/σ_y, t_j/σ_t,) where position values are given by the values (x, y, in μm) of the channel, n, on which the proto-event was detected, t is the time of the proto-event (in ms), and σ_x, σ_y and σ_t define distance scales in space and time. Gradient-ascent clustering (GAC, Swindale and Spacek 2014) which is based on the mean-shift procedure (Fukunaga and Hostetler 1975) was used to cluster the proto-events. The points, $\mathbf{p}_j$ are first duplicated to form a second set of points (termed 'scout' points) $\mathbf{s}_k$, k=1, K; K=J initially. At each step, each scout point $\mathbf{s}_k$ calculates a local gradient of attraction from the values of the nearby points in $\mathbf{p}$. In this particular implementation of the algorithm the gradient was weighted by the absolute value of the channel voltage, V_k, of the corresponding proto-event. Scout points then move in space and time up the gradient. This tends to move points belonging to the same event closer together and towards the local voltage peak of the waveform. The distances between pairs of scout points are then calculated and points that are closer than a threshold distance, d_m, are merged, thereby shortening the list. Note that the list of points, $\mathbf{p}$, from which the gradients are calculated remains unchanged throughout the clustering procedure. A new set of gradients, as seen by the new positions of scout points $\mathbf{s}$, is then calculated and scout points move and are merged as before. These steps are repeated for a sufficiently large number of iterations, or until the number of clusters, K, stops changing and/or the movement of scout points falls below a threshold value. The shortened list of scout points, $\mathbf{s}_k$, is now the list of events.

More specifically, at each iteration each scout point is moved by an amount

$$\Delta\mathbf{s}_k = \frac{\sum_{j=1}^{J} V_j\left(\mathbf{p}_j - \mathbf{s}_k\right) G\left(\mathbf{p}_j, \mathbf{s}_k\right)}{\sum_{j=1}^{J} V_j G\left(\mathbf{p}_j, \mathbf{s}_k\right)} \tag{4}$$

where $G\left(\mathbf{p}_j, \mathbf{s}_k\right) = e^{-d^2}$ is a Gaussian function of the distance, d, between points $\mathbf{p}_j$ and $\mathbf{s}_k$, with $d(j,k)$ defined as:

$$d(j,k) = \left[\left(\frac{x_j - x_k}{\sigma_x}\right)^2 + \left(\frac{y_j - y_k}{\sigma_y}\right)^2 + \left(\frac{t_j - t_k}{\sigma_t}\right)^2\right]^{\frac{1}{2}}. \tag{5}$$

Scout points were merged if they were less than a distance d_m =0.25 apart.

The summation in Eq. (4) is potentially slow if there are many proto-events. However there are several ways of speeding it up. Proto-events are indexed both by channel number and by time. If they are stored in temporal order (e.g. in an array where channel number varies most rapidly) the summation can be done by searching locally within the array since those points that are far apart in the array will be sufficiently far apart in time that their contribution to the gradient can be ignored. This makes execution time vary linearly with recording duration. The calculation can be further sped up (several-fold) in the following ways:

(a) the indices of the neighbors of each proto-event that are less than a distance $4\sigma_x$ apart can be tabulated, avoiding terms in Eq. (4) that are likely to be close to zero. This requires storing, for each proto-event, a list of the indices of the other nearby proto-events, plus an index into this table and the number of entries in it;
(b) proto-events that have no neighbors can be tagged as isolated and omitted from the subsequent clustering process;
(c) the calculation of distances between points in $\mathbf{s}$ for merging can likewise be sped up by searching locally within the array for neighbors;
(d) merging pairs of points in $\mathbf{s}$ can be simplified by simply deleting one of them as there is no need to keep track of all the proto-events that belong to one cluster. This can be done by setting the entry in the table of indices to zero,

and skipping these points on subsequent iterations. The remaining points in **s** remain in the same position in the table, i.e. in alignment with those in **p**.

The algorithm requires choices for the values of the distance scales σ_x, σ_y and σ_t. If these values are too small, single spikes may give rise to duplicate events; if they are too large, spikes that are close together in space and time will not be resolved. We normally set $\sigma_y \equiv \sigma_x$ and examined the effects of varying σ_x and σ_t on event duplication and event resolution.

2.5 Code

Program code was written using Intel Visual Fortran Composer XE running under Windows 7 (64 bit) on an Intel Sandy Bridge Core i7 CPU at 3.4 GHz. The entire voltage record was read from disk and stored in RAM. Parallel optimizations were not used for any of the procedures.

3 Results

3.1 Simulated noise

Figure 1A shows that the autocorrelation functions of periods of spike-free noise in real data, and of simulated noise, were similar. Visually, the simulated recordings looked similar to real ones (Fig. 1B).

3.2 Comparison of detection algorithms

We compared the ability of 3 different procedures to detect small amplitude spikes: the bipolar threshold (BP) method, the non-linear energy operator (NEO) and the dynamic multiphasic (DMP) detector. The DMP detector requires specification of a temporal window, δ, and we tested values of δ =0.16, 0.24, 0.32 and 0.4 ms. The methods were tested with spike waveforms with a scaled peak to peak amplitude of 80 μV added at known times to simulated noise with a standard deviation of 10 μV. Because the spikes were far apart in time it made no difference whether the STL or PEC methods for avoiding duplication were used. Detection was run with a series of different thresholds and, for each threshold, we calculated the percentage of events that was correctly detected (% correct, or true positive rate) and the false positive (FP) rate. Because the total number of false positives will scale with the number of channels and the duration of the recording, the total number cannot be directly compared with the percent correct. Hence standard ROC curves cannot be constructed. Instead we expressed the FP rate as the rate per channel per minute. This rate and percent correct will both be high for low thresholds and both will decrease as the threshold increases. The better the detection method, the lower the false positive rate will be for a given detection accuracy, hence the closer the curve is to the bottom right-hand corner of the graph, the better the method. The results (Fig. 4) showed that the DMP detector generally produced roughly an order of magnitude fewer false positives than either the NEO or BPT methods. Only in one case (S1) did the NEO detector perform as well as the best of the other methods; in all other cases the NEO was worst. The choice of temporal window also had a significant effect: performance was generally best with δ =0.24 ms or longer. A short window, δ =0.12 ms, performed worse for long duration spikes (S3 and S6); for the other 4 test spikes, window duration had a much smaller effect. The relative advantage of the DMP method was greater in spikes where the positive and negative phases were of roughly equal amplitude (S2, S4 and S6). Given these results, the DMP method with δ =0.24 ms was chosen for all subsequent tests.

We next examined the effect of spike amplitude on detection accuracy. This was done by creating simulated recordings 5 min long, to which 1000 spikes scaled to a given fixed peak to peak voltage were added at known times. We calculated the percentage of correctly identified events as a function of peak-to-peak amplitude for each of the test spike waveforms (Fig. 5). For these tests we used the DMP method with a threshold θ =3.4×noise and δ =0.24 ms. This resulted in a roughly constant false positive rate of 17 events/min/channel. Accuracy was >99 % for spikes with a peak-to-peak amplitude>than 100 μV. Spikes with spatially extensive waveforms (e.g. S2, S3 and S5) were better detected than spikes with narrower distributions (e.g. S1 and S4).

3.3 Avoiding duplicate events

Application of the DMP detection method (or of any similar method) will often result in several threshold crossings per spike waveform. For example, Fig. 3 shows a spike giving rise to two detected events on a single channel. This will be especially likely for spikes that are spatially and/or temporally extensive, and for spikes of large amplitude. We compared the ability of two different methods, spatiotemporal lockout (STL) and proto-event clustering (PEC) to avoid this problem. Both methods require the definition of spatiotemporal window parameters: L_x and L_t for the STL method, and σ_x and σ_t for the PEC method. It should be noted that because L_x defines a discrete window and because channels are discretely spaced, small changes in L_x often have no effect. We asked what were the minimum window sizes required to prevent event duplication – the most common outcome if window sizes are poorly chosen. To do this, a second set of 6 test spike waveforms (Fig. 2; S7 – S12) was selected as examples of spikes that seemed particularly likely to give rise to duplicates. These included spikes that were spatially and/or temporally extended and of relatively large amplitude. In each simulated recording,

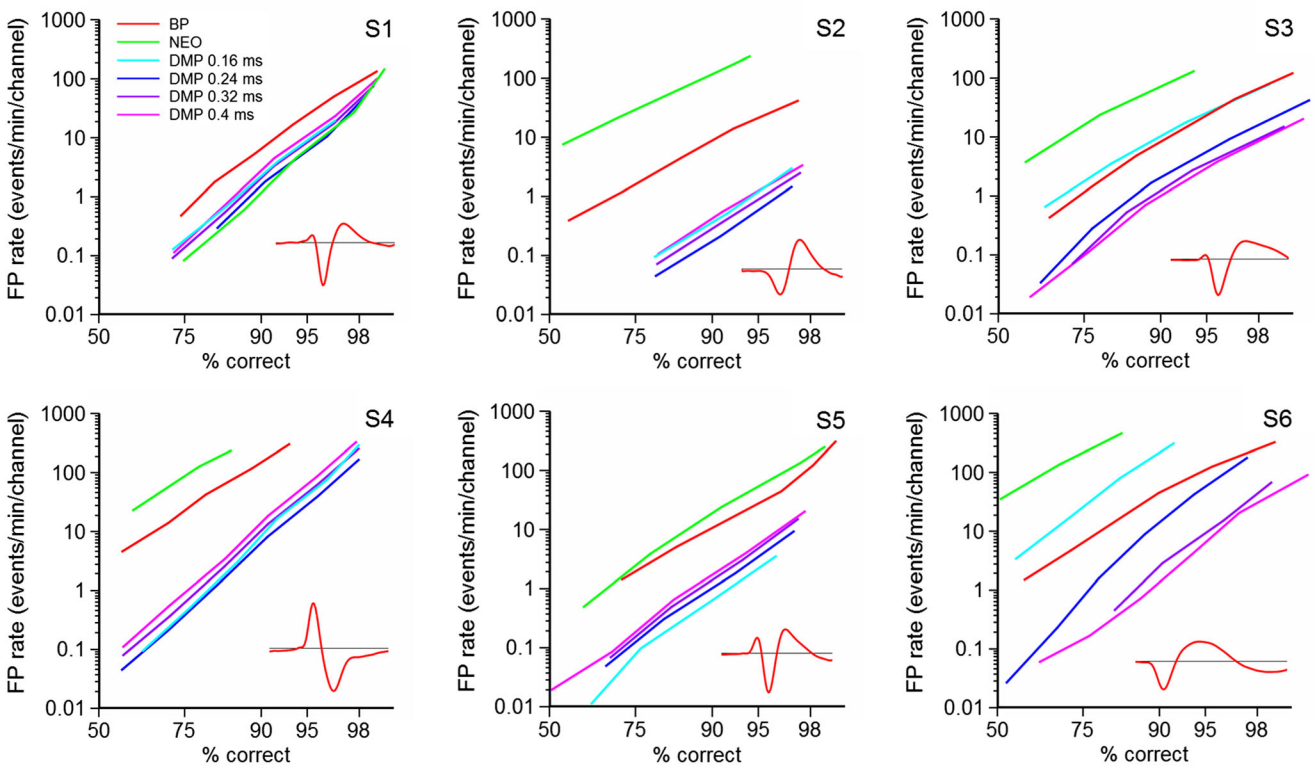

Fig. 4 Plots of false positive (FP) rate (events/min/channel) vs. detection accuracy (% correct) for spikes S1 – S6 (Fig. 2), for a variety of event detection methods: BP – bipolar threshold; NEO – non-linear energy; DMP – dynamic multiphasic with temporal windows δ as specified. Each curve is derived from a single simulated recording 5 min long to which duration 1000 of the test spikes shown in Fig. 2, scaled to a peak-to-peak amplitude of 80 μV, were added at known times. Detection thresholds were varied to produce different detection accuracies and false positive rates. An event was identified as correct if it occurred within ±0.5 ms and <250 μm of the centre channel of the test spike. All other events were counted as false positives. The probability of a false positive being wrongly identified as a real event was $<10^{-6}$. Probability scaling is used on the horizontal axis and logarithmic scaling on the vertical axis

1000 test waveforms from one member of the set were added to noise, without voltage scaling, at 60 ms intervals. Event detection was run on each of these 6 simulations, using the DMP method with a threshold θ =4.0 and δ =0.24 ms. This usually resulted in the detection of >99 % of the events and a false positive rate of 0.26 events/min/channel. The number of duplicate events was calculated by counting pairs of events that were within ±2.5 ms and less than 250 μm of each other. Duplicates can be of two kinds: temporal or spatial. Temporal duplicates typically occur on the center channel a short time (<1 ms) after the main event, usually because of a slow negative after potential. Spatial duplicates occur when a spike waveform crosses threshold on a channel some distance from the main one, usually at nearly the same time as the other event. We examined the effects of changing L_x and σ_x on the frequency of spatial duplicates (setting L_t and σ_t to values for which temporal duplicates did not occur) and, conversely, we examined the effects of changing L_t and σ_t on the frequency of temporal duplicates (setting L_x and σ_x to values for which spatial duplicates did not occur). Figure 6A compares the number of temporal duplicates produced by the STL and PEC methods as a function of L_t or σ_t, for each of the 6 test spike templates. STL gave somewhat variable results with the different spikes, requiring values of $L_t \geq 0.6$ ms to avoid rates of event duplication >0.5 % on all six test waveforms. PEC consistently avoided duplications with $\sigma_t \geq 0.25$ ms. A similar pattern of results was seen for the effects of L_x and σ_x on spatial duplication (Fig. 6B): here a value of $L_x \geq 150$ μm was required to keep the duplicate rate below 0.5 % in all six test waveforms, while for PEC, a value of $\sigma_x \geq 80$ μm was sufficient.

Using these results as a guide, values of L_x =150 μm, L_t =0.6 ms, σ_x =80 μm and σ_t =0.25 ms were chosen for subsequent tests of spatiotemporal resolution.

3.4 Spatiotemporal resolution of STL vs PEC

We asked what are the smallest spatial and temporal distances between a pair of spikes that will still allow both of them to be resolved by event detection? To measure this, pairs of spike waveforms from S7 – S12, were added to simulated noise with varying, known, spatial and temporal offsets. We then ran event detection using either the STL or PEC methods and calculated, for each spatial and temporal offset, the fraction of spike pairs where two events were detected at the known locations of the added waveforms. Figure 7 shows that both

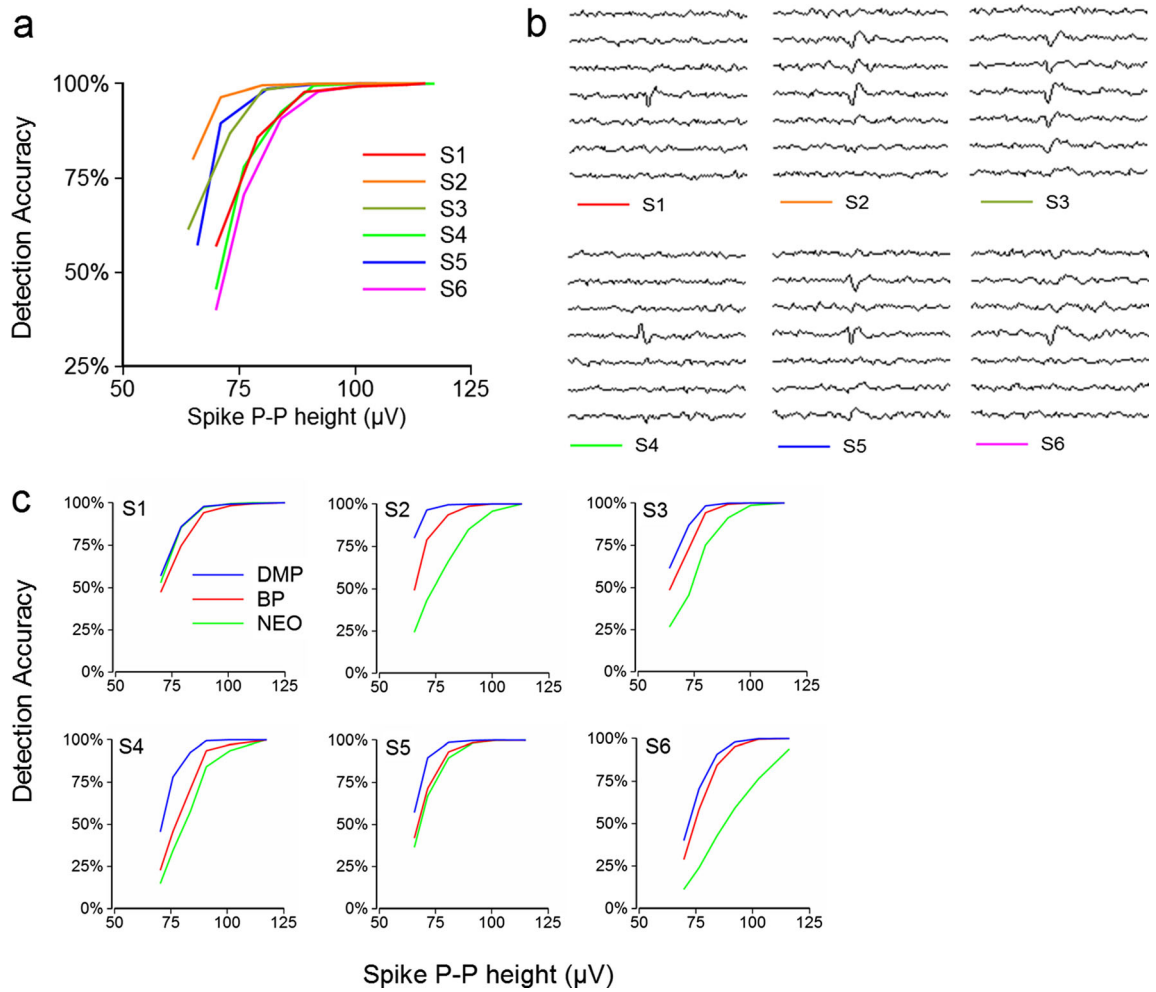

Fig. 5 **a** Detection accuracy as a function of spike peak-to-peak height for spikes S1 – S6, measured with the DMP method with a temporal window of 0.24 ms. **b** Examples of test spikes S1 – S6 with a peak-to-peak height scaled to 100 μV in simulated noise. Vertical layout of channels corresponds with that of the layout shown in Fig. 2, with the channels taken in reading order. **c** Comparison of the detection accuracy of the DMP, BP and NEO methods as a function of spike peak-to-peak height. A fixed threshold was used for each method, adjusted to give the same false positive rate (17.15 events per minute per channel) for each method

methods could resolve spikes as close as 0.8 ms apart, regardless of spatial separation. Likewise, spikes more than 200 μm apart could be resolved, regardless of temporal separation.

3.5 Processing time

We measured the time taken to detect events using a set of 15 real recordings of varying length. Despite the greater algorithmic complexity, event detection times for PEC were similar to, or in some cases faster than, those for STL. In general, detection time was roughly 50 times faster than recording duration. The time taken to detect events was strongly correlated with recording duration (Fig. 8A) and more weakly correlated with the number of events detected (Fig. 8B). The relatively weaker correlation between the number of events and detection time suggests that a significant fraction of both algorithms' computation time was devoted to scanning regions of the data file lacking detectable events.

4 Discussion

Spike detection with polytrodes and MEAs has a variety of problems, some common to all types of microelectrode recording, and some specific to electrodes designed to resolve many different units at once. Spikes can fail to be detected, either because they are of low amplitude, or because the detection method is insensitive to particular kinds of shape. False positives may be present. Problems specific to polytrodes and MEAs include duplication of events, either

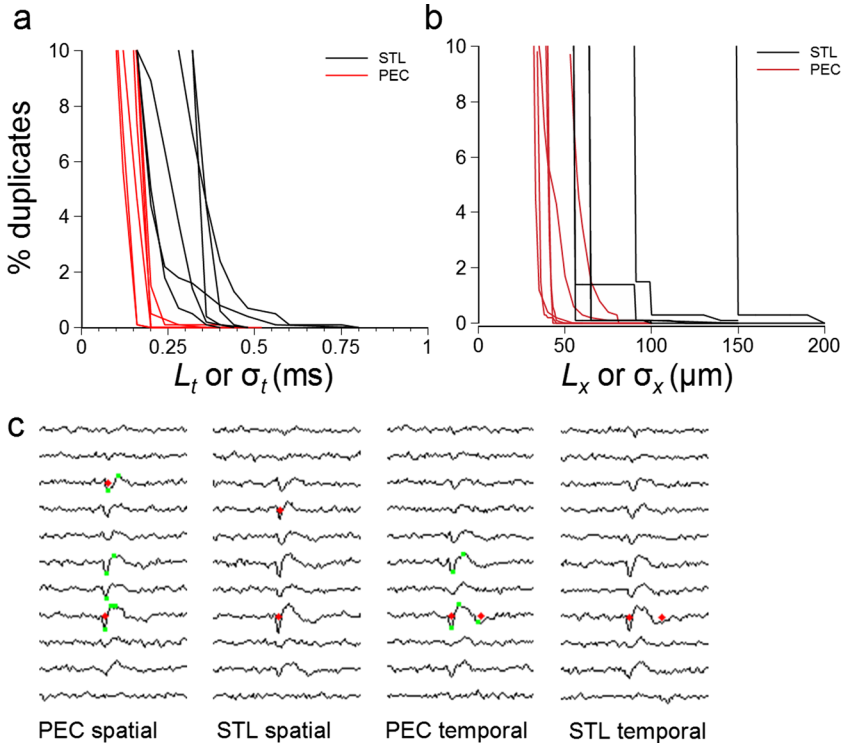

Fig. 6 **a** The percentage of temporal duplicate events resulting from different choices of the temporal windows L_t, (STL *black lines*) or σ_t (PEC, *red lines*) for test sets S7 – S12. **b** Percentages of spatial duplicates as a function of the spatial windows L_x (STL, *black lines*) and σ_x (PEC, *red lines*). Each trace is from a single test set. Note that the PEC method more reliably reduces the percentage of duplicates to near zero as the window size increases. **c** Examples of duplicates produced by the different methods resulting from too small a value of either the spatial (σ_x =50 μm or L_x =75 μm) or the temporal (σ_t =0.2 ms or L_t =0.4 ms) window. Examples come from test set S12. *Red dots* indicate the channel to which the event was localised; *green dots* (PEC method only) show the locations of the proto-events

spatially or temporally. Spikes that occur close together in space and time may also fail to be resolved. In this paper we examined the behaviour of several detection methods with the goal of reducing these problems. We showed that, out of 3 methods tested, one, the DMP method, which involves thresholding plus subtraction of voltages at time points a short interval apart, performs significantly better than either simple bipolar thresholding or energy detection, giving rise to a higher proportion of correctly detected events with fewer false positives for a variety of spike shapes (Figs. 4 & 5). Bipolar thresholding and energy detection are both currently popular methods of event detection used by commercially available spikesorting packages (Maccione et al. 2009). Of the two methods for avoiding event duplication, STL and PEC, the latter method offered a more reliable relation between the choice of spatial and temporal window parameters and the avoidance of duplicates. Both methods offered similar spatio-temporal resolution: only spike pairs that were <200 μm apart in space and <0.8 ms apart in time could not be resolved. Finally, on our 54 channel polytrode recordings, the CPU time taken for event detection scaled linearly with recording duration and was roughly 50 times faster than real time for both methods. We discuss these results in more detail as follows.

4.1 Initial detection of events

The dynamic multiphasic event detector (DMP) performed significantly better than a simple bipolar threshold (BP) or the non-linear energy operator (NEO) of Kaiser (1990). For a given percentage of correctly detected spikes, false positive rates were, for 5 out of 6 of our test spike shapes, an order of magnitude smaller than with the other two methods. For the 6th spike (S1, Fig. 2) the DMP detector's performance was similar to that of the NEO. It is likely that the NEO did well with this particular spike because of the sharpness of its initial negative peak. We did not evaluate the performance of the hyper-ellipsoidal thresholding algorithm of Rebrik et al. (1999), because earlier tests (Blanche 2005) suggested that its performance was comparable to, or worse, than that of the BP and NEO detectors. This algorithm deliberately discounts noise correlations between different channels, however since such correlations can be caused by spike waveforms it would seem more useful to attempt to preserve rather than discard them. It is also computationally more complex than the other methods (Blanche 2005).

Our implementation of the DMP method is based on that described in Blanche (2005) and is similar to recent proposals (Borghi et al. 2007; Maccione et al. 2009). However a

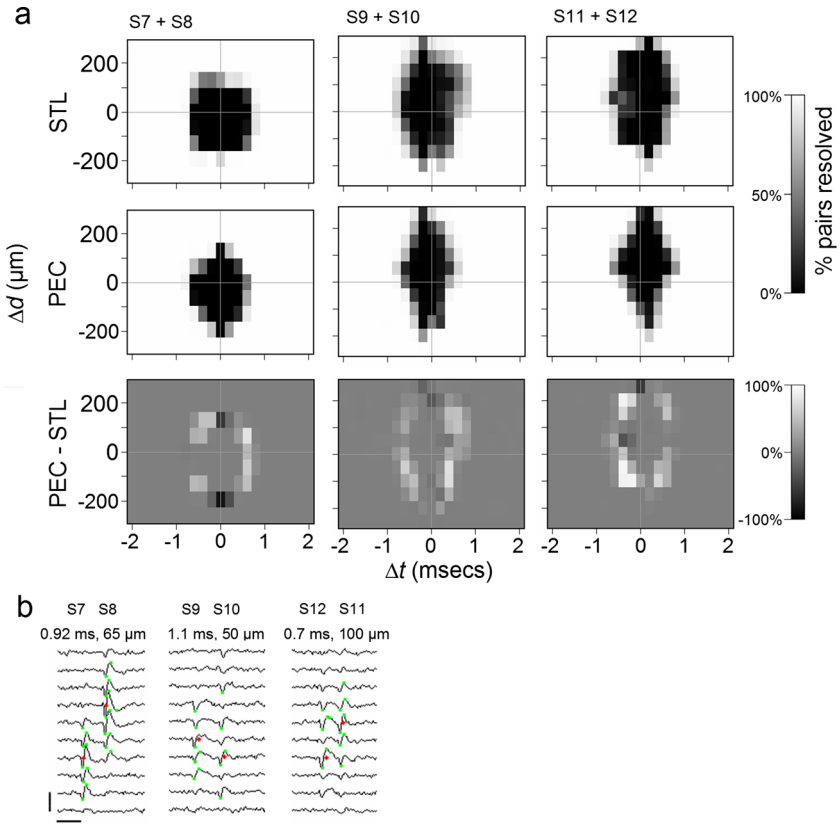

Fig. 7 **a** *Upper two rows*: the spatiotemporal resolution of the STL and PEC methods, determined by running event detection with pairs of spikes from set S7 – S12, separated by particular spatial (Δd) and temporal (Δt) offsets. *Lower row*: the difference between PEC and STL methods (middle row minus the top row). **b** Examples of pairs of spikes resolved with the PEC method that would be less likely to be resolved with the STL method. The spike sets, and temporal and spatial separation are shown above each panel. *Red dots* indicate the event location and *green dots* indicate the locations of the proto-events. *Scale bars* show 1 ms and 200 μV. Note that the physical distance between the spike center channels (*red dots*) does not match the distance on the page because the channel layout on the electrode only approximately matches the layout on the page

difference between our method and these latter proposals is that we do not require more than one local maximum (or minimum) to be present. This is important, because noise can make individual spike waveforms unipolar (i.e. possess only a single local inflection point) even if, on average, more than one inflection point is present. In initial tests we found that comparing the differences between successive minima and maxima resulted in the loss of some events from nearly unipolar spikes because of the absence of a second inflection point within the time window. This consideration seems especially relevant given the possibility that unfiltered spike shapes may be unipolar, or more nearly so, than in filtered recording data (Wiltschko et al. 2008; Quian Quiroga 2009).

A possible reason for the superiority of the DMP detector is that it involves a subtraction of two voltage values at a time interval of around 0.25–0.5 ms. If the noise in these values is uncorrelated, the noise in the estimate of the voltage difference will be reduced by $\sqrt{2}$ and hence the signal-to-noise ratio will increase by that amount. Our measurements showed that voltage values are weakly (negatively) correlated in this time interval (Fig. 1a). Detection might be further improved if it was possible to take 3 weakly correlated samples into account, however many spikes are only long enough to allow for two such samples. Variations in the choice of δ might be used to bias detection towards fast or slow

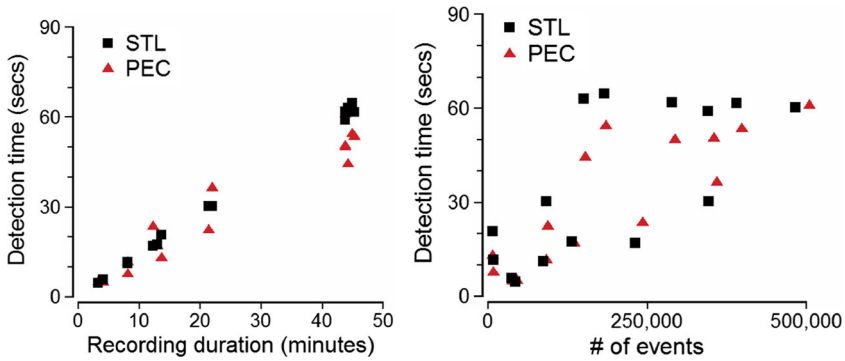

Fig. 8 Time taken for event detection correlates strongly with the duration of the recording (*left panel*; r=0.998 (STL) and 0.97 (PEC)) and, less strongly, with the total number of events (*right panel*; r=0.70 (STL) and 0.85 (PEC)). Times are from 15 different recordings (of non-simulated data) of varying length and number of events

waveforms, e.g. smaller values might lead to a reduction in false positives if detection was purposely biased to fast spiking neurons. It should be noted that noise correlations may differ depending on the characteristics of the recording hardware and brain region and species being studied. These differences might affect the relative performances of the different methods.

Noise values in our recordings were weakly correlated across nearby channels (data not shown). We did not include such correlations in our simulated noise on the grounds that detection *per se* of the DMP, BP and NEO methods relies solely on within-channel voltage measurements. Cross-channel noise correlations would increase the probability of low amplitude events and proto-events being detected simultaneously on neighbouring channels (as would the correlations induced by spikes) and might therefore impact the occurrence of duplicates. Proto-event clustering and possibly estimates of spatio-temporal resolution might also be affected. Further simulations would probably be necessary to determine whether this would make the performace of the algorithms better or worse.

Although the ability to detect spikes with very small signal-to-noise ratios may seem important, there is a caveat if the spikes are going to be sorted subsequently. The differences in spike shape between physiologically distinct units can be small (Swindale and Spacek 2014), and in the inevitable presence of noise, small spikes from different units may be indistinguishable. Small amplitude events are often very numerous (possibly because they consist of many different units). Hence, while using a low detection threshold and an efficient detection method may result in the detection of many genuine low-amplitude spikes, these may be impossible to sort reliably later on. If there is a strong requirement for well-sorted units it may be better to choose a higher detection threshold and present the sorter with a smaller number of relatively easy to sort waveforms. A related consideration is that it may be incorrect to consider that 'noise' is a background signal which has a different origin from that of the signal (spikes) which is to be detected. Unfiltered electrical signals in the brain are the sum of currents resulting from synaptic currents as well as action potentials (Buzsáki et al. 2012). Filtering to remove low frequencies (the LFP component) may preferentially reduce the synaptic contribution since EPSPs and IPSPs are generally slower than action potentials. If the 'noise' seen in high-pass filtered recordings used for spike sorting is largely the sum of large numbers of unresolvable, low-amplitude spikes then there is unlikely to be a detection method that can cleanly separate signals from noise. In this case, event detection becomes merely a pragmatic method for choosing a subset of signals likely to be amenable to classification at a subsequent stage. Whether this is true or not, in practice, whatever detection threshold is chosen, there will almost always be some units for which a of fraction spikes fall below threshold and is lost. This can be shown by looking for sharp boundaries in the distribution of peak-to-peak height in the waveforms of sorted units. A judgement might then be made as to how big the fraction is and whether the unit should be deleted or not.

4.2 Template-matching methods

A different class of detection methods works by matching predefined spike templates to recorded signals (e.g. Segev et al. 2004; Thakur et al. 2007; Franke et al. 2010; Pillow et al. 2013). Since the matching can be performed with multichannel templates, issues of duplication do not arise. In addition, such methods can be adapted to resolve temporally and spatially overlapping spikes. These advantages notwithstanding, template-matching methods have significant disadvantages especially when applied to cortical recordings. Spike templates have first to be defined by identifying and sorting spikes from an initial sample period of recording. Subsequently, the method will either fail to identify units that did not fire during the sample period or will mis-classify them. Hence the method is not well suited to identifying units that fire sporadically or at low rates or which are driven only by very specific kinds of stimuli. This seems a major disadvantage for cortical recordings where such behaviours are expected (Attwell and Laughlin 2001; Lennie 2003; Mizuseki and Buzsáki 2013). Another disadvantage is that matching is often done using the rms difference between the template and the candidate spike as a criterion for identification. However the rms differences between different units can be very small even though principal components analysis and physiological receptive field properties indicate that the units are different (Swindale and Spacek 2014). For these reasons we did not attempt to apply or evaluate detection methods based on template-matching.

4.3 Duplicate events

Duplicate events – the same spike being counted twice - can be avoided entirely by use of a sufficiently large spatial and temporal exclusion zone around each detected event. However this results in a loss of spatiotemporal resolution. On polytrodes and MEAs, different units can be very close to each other and their spikes might occur at any temporal interval so it is desirable to make the procedure for excluding duplicates as efficient as possible. Duplicates can be classified into two kinds: spatial and temporal (intermediates are possible, but were rarely observed). The consequences of each kind for subsequent sorting are different. Spatial duplicates are events that occur at almost the same time (e.g. <0.24 ms), on different channels and which belong to the same spike. Clustering of these events will normally result in two clusters

where the spikes in the smaller cluster will be a subset of those in the other (assuming that not all the spikes gave rise to duplicate events). If only a small number of duplicates occurs, the cluster may fall below a threshold size and be deleted, thereby correcting the problem. If this does not happen, the two clusters may be recognized as similar on the basis of subsequent pairwise comparisons of their multichannel waveforms, and combined (Swindale and Spacek 2014). Once this is done, the cluster can be scanned for the presence of spike pairs that are simultaneous, or very nearly so, and one member of each pair can be deleted. Such tests can be done automatically following alignment of events in a cluster, and can be fast and reliable. Hence spatial duplicates do not pose a severe problem for spike sorting. Temporal duplicates can be more difficult. It is possible that, at some stage after detection, the event will be realigned (e.g. to a dominant peak in the waveform) and this may move it so that the time of occurrence becomes identical to the other event. If this happens, the event can be deleted by checks for event simultaneity, as with spatial duplicates. Otherwise, the duplicate events may show up as a cluster and it may take manual inspection to determine that it is spurious. Although manual inspection will be a normal part of most spike sorting procedures, it is preferable if problems like this can be avoided at an early stage. Alternatively, the presence of a very strong cross-correlation between the two units might be used to flag that one of them is spurious.

4.4 Spatiotemporal resolution

Methods for resolving spike waveforms that overlap in time and space have been proposed (Lewicki 1994; Chandra and Optican 1997; Lewicki 1998; Sahani et al. 1998; Segev et al. 2004; Zhang et al. 2004; Ding and Yuan 2008; Franke et al. 2010; Pillow et al. 2013). These procedures are necessary when nearby cells fire at relatively high rates, as in the retina, when a large fraction of waveforms can be expected to overlap. Such methods are computationally intensive but have the advantage that spike train cross-correlograms of spatially adjacent neurons can be computed more accurately for short time intervals. Here, we have attempted to establish the limits of spatiotemporal resolution that can be achieved without such techniques. This can be justified for cortical recordings on the grounds that there is both theoretical (Attwell and Laughlin 2001; Lennie 2003) and experimental evidence (Brecht et al. 2003; Mizuseki and Buzsáki 2013; Wohrer et al. 2013) that cortical neurons fire, on average, at rates below 1 Hz. Thus temporally overlapping pairs are expected to be relatively infrequent. This argument does not apply in cases, such as the retina, where nearby cells fire simultaneously at high rates. Overlapping spikes may be more common during such periods. We found that for units closer than 200 μm, there is a temporal 'black hole' of about ±0.8 ms within which spike pairs cannot be resolved (Fig. 7). We have argued elsewhere (Swindale and Spacek 2014) that for a typical cortical cross-correlogram with a binwidth of 10 ms (or more) only 16 % of spikes (or less) would be missing from the central bin. The percentage will be higher, of course, if more highly resolved cross-correlations are being studied. However, use of the present methods of event detection followed by sorting does not rule out the use of a subsequent stage in which attempts are made to find and resolve waveforms into pairs that can be assigned to existing clusters.

4.5 PEC vs STL

Proto-event clustering (PEC) had advantages over spatiotemporal lockout (STL), though these advantages were relatively small. PEC seemed more reliable as a method for avoiding event duplication (Fig. 6), offered slightly better spatiotemporal resolution (Fig. 7a) and was slightly faster (Fig. 8). Conceptually, clustering seems a more principled approach to the problem than simply applying a fixed window following the initial detection of an event. Though not examined here, centering the lockout region on the spike following initial detection can sometimes be problematic and the PEC method avoids this. Given that its job is to find and resolve small clusters in space-time, the PEC method might also be improved, at the expense of speed, by changing the detection method to yield a higher number of proto-events per spike, e.g. by including all threshold crossings and not just local maxima and minima.

Acknowledgments This work was supported by grants from the Canadian Institutes of Health Research and the Natural Sciences and Engineering Research Council of Canada.

Conflict of interest The authors declare that they have no conflict of interest.

References

Attwell, D., & Laughlin, S. B. (2001). An energy budget for signalling in the grey matter of the brain. *Journal of Cerebral Blood Flow and Metabolism, 21*, 1133–1145.

Blanche, T. J. (2005). *Large Scale Neuronal Recording*. Ph. D. Thesis, University of British Columbia.

Blanche, T. J., Spacek, M. A., Hetke, J. F., & Swindale, N. V. (2005). Polytrodes: high density silicon electrode arrays for large scale multiunit recording. *Journal of Neurophysiology, 93*, 2987–3000.

Borghi, T., Gusmeroli, R., Spinelli, A. S., & Baranauskas, G. (2007). A simple method for efficient spike detection in multiunit recordings. *Journal of Neuroscience Methods, 163*, 176–180.

Bragin, A., Hetke, J., Wilson, C. L., Anderson, D. J., Engel, J., Jr., & Buzsáki, G. (2000). Multiple site silicon-based probes for chronic recordings in freely moving rats: implantation, recording and histological verification. *Journal of Neuroscience Methods, 98*, 77–82.

Brecht, M., Roth, A., & Sakmann, B. (2003). Dynamic receptive fields of reconstructed pyramidal cells in layers 3 and 2 of rat somatosensory barrel cortex. *Journal of Physiology, 553*, 243–265. doi:10.1113/jphysiol.2003.044222.

Buzsáki, G. (2004). Large-scale recording of neuronal ensembles. *Nature Neuroscience, 7*, 446–451.

Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents – EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience, 13*, 407–420.

Chandra, R., & Optican, L. M. (1997). Detection, classification, and superposition resolution of action potentials in multiunit single-channel recordings by an on-line real-time neural network. *IEEE Transactions on Biomedical Engineering, 44*(5), 403–412. doi:10.1109/10.568916.

Choi, J. H., Jung, H. K., & Kim, T. (2006). A new action potential detector using the MTEO and its effects on spike sorting systems at low signal-to-noise ratios. *IEEE Transactions on Biomedical Engineering, 53*(4), 738–746.

Ding, W., & Yuan, J. (2008). Spike sorting based on multi-class support vector machine with superposition resolution. *Medical and Biological Engineering and Computing, 46*(2), 139–145. doi:10.1007/s11517-007-0248-0.

Franke, F., Natora, M., Boucsein, C., Munk, M., & Obermayer, K. (2010). An online spike detection and spike classification algorithm capable of instantaneous resolution of overlapping spikes. *Journal of Computational Neuroscience, 29*, 127–148.

Fukunaga, K., & Hostetler, L. D. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory (IEEE), 21*, 32–40.

Gold, C., Henze, D. A., Koch, C., & Buzsáki, G. (2006). On the origin of the extracellular action potential waveform. *Journal of Neurophysiology, 95*, 3113–3128.

Jäckel, D., Frey, U., Fiscella, M., Franke, F., & Hierlemann, A. (2012). Applicability of independent component analysis on high-density microelectrode array recordings. *Journal of Neurophysiology, 108*, 334–348.

Kaiser, J. F. (1990). On a simple algorithm to calculate the 'energy' of a signal. *Proceedings of the ICASSP, 1*, 381–384.

Kim, K. H., & Kim, S. J. (2000). Neural spike sorting under nearly 0-dB signal-to-noise ratio using nonlinear energy operator and artificial neural-network classifier. *IEEE Transactions on Biomedical Engineering, 47*, 1406–1411.

Lennie, P. (2003). The cost of cortical computation. *Current Biology, 13*, 493–497. doi:10.1016/S0960-9822(03)00135-0.

Lewicki, M. (1994). Bayesian modelling and classification of neural signals. *Neural Computation, 6*(5), 1005–1030. doi:10.1162/neco.1994.6.5.1005.

Lewicki, M. (1998). A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems, 9*(4), R53–R78.

Maccione, A., Gandolfo, M., Massobrio, P., Novellino, A., Martinoia, S., & Chiappalone, M. (2009). A novel algorithm for precise identification of spikes in extracellularly recorded neuronal signals. *Journal of Neuroscience Methods, 177*, 241–249.

Maragos, P., Kaiser, J. F., & Quatieri, T. F. (1993). On amplitude and frequency demodulation using energy operators. *IEEE Transactions on Signal Processing, 41*, 1532–1550.

Mizuseki, K., & Buzsáki, G. (2013). Preconfigured, skewed distribution of firing rates in the hippocampus and entorhinal cortex. *Cell Reports, 4*(5), 1010–1021.

Obeid, I., & Wolf, P. D. (2004). Evaluation of spike-detection algorithms for a brain-machine interface application. *IEEE Transactions on Biomedical Engineering, 51*(6), 905–911. doi:10.1109/TBME.2004.826683.

Pillow, J. W., Shlens J., Chichilnisky, E. J., & Simoncelli, E. (2013). A model-based spike sorting algorithm for removing correlation artifacts in multi-neuron recordings. *PLoS One*, *8*(5). doi:10.1371/journal.pone.0062123.

Quian Quiroga, R. (2009). What is the real shape of extracellular spikes. *Journal of Neuroscience Methods, 177*, 194–198.

Quian Quiroga, R., Nadasdy, Z., & Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation, 16*, 1661–1687.

Rebrik, S., Wright, B., Emondi, A., & Miller, K. (1999). Cross-channel correlations in tetrode recordings: implications for spike-sorting. *Neurocomputing, 26*, 1033–1038.

Sahani, M., Pezaris, J. S. & Andersen, R. A. (1998). Extracellular recording from multiple neighboring cells: a maximum-likelihood solution to the spike-separation problem. In: *Computational Neuroscience,* edited by J. Bower, Plenum Press, New York, pp. 619–625.

Segev, R., Goodhouse, J., Puchalla, J., & Berry, M. J. (2004). Recording spikes from a large fraction of the ganglion cells in a retinal patch. *Nature Neuroscience, 7*, 1155–1162. doi:10.1038/nn1323.

Swindale, N. V., & Spacek, M. A. (2014). Spike sorting for polytrodes: a divide and conquer approach. *Frontiers in Systems Neuroscience, 8*(6), 1–21. doi:10.3389/fnsys.2014.00006.

Thakur, P. H., Lu, H., Hsiao, S. S., & Johnson, K. O. (2007). Automated optimal detection and classification of neural action potentials in extra-cellular recordings. *Journal of Neuroscience Methods, 162*, 364–376. doi:10.1016/j.jneumeth.2007.01.023.

Wiltschko, A. B., Gage, G. J., & Berke, J. D. (2008). Wavelet filtering before spike detection preserves waveform shape and enhances single-unit discrimination. *Journal of Neuroscience Methods, 173*, 34–40. doi:10.1016/j.jneumeth.2008.05.016.

Wohrer, A., Humphries, M. D., & Machens, C. K. (2013). Population-wide distributions of neural activity during perceptual decision-making. *Progress in Neurobiology, 103*, 156–193. doi:10.1016/j.pneurobio.2012.09.004.

Zhang, P. M., Wu, J. Y., Zhou, Y., Liang, P. J., & Yuan, J. Q. (2004). Spike sorting based on automatic template reconstruction with a partial solution to the overlapping problem. *Journal of Neuroscience Methods, 135*(1–2), 55–65. doi:10.1016/j.jneumeth.2003.12.001.