Phased-Array Processing for Spike Discrimination

Yikun Huang and John P. Miller
Center for Computational Biology, Montana State University, Bozeman, Montana 59717

Submitted 29 January 2004; accepted in final form 10 April 2004

Huang, Yikun and John P. Miller. Phased-array processing for spike discrimination. J Neurophysiol 92: 1944–1957, 2004. First published April 28, 2004; 10.1152/jn.00088.2004. We present a novel approach for the detection, discrimination, and identification of superimposed neuronal action potentials from multineuronal, multichannel extracellular nerve recordings with low signal-to-noise ratios. The approach uses phased-array processing techniques to identify the spikes from different neurons on the basis of their unique propagation velocities. We evaluated this new approach using simulated electrophysiological data, under conditions that are known to limit the effectiveness of existing spike discrimination techniques. This approach enabled discrimination of simulated spikes from multiple simultaneously active neurons with a high degree of reliability and robustness within the expected range of experimental recording conditions, even in situations where there was a high degree of spike waveform superposition on the recording channels. Moreover, the technique enables the reliable detection and discrimination of spikes recorded with signal-to-noise ratios less than 1.

INTRODUCTION

A detailed understanding of neural coding and distributed processing requires the characterization of the simultaneous activity of large ensembles of neurons, and neurophysiological research continues to drive the development of new tools and approaches for multineuron recording and analysis. The first and most problematic stage in analyzing extracellular neuronal electrophysiological recordings is spike discrimination, that is, the decomposition of the multineuronal composite voltage waveforms recorded with extracellular electrodes into sequences of individually resolved action potentials assignable to distinct neurons (or “units”). Several factors make multunit spike discrimination analysis difficult and potentially unreliable. In particular, it is extremely difficult to discriminate action potentials from different neurons when those spikes are partially or totally superimposed on the recording channels. We present a novel approach for discriminating and identifying superimposed spikes from unit extracellular nerve recordings.

This new approach uses phased-array analysis techniques, and offers several significant advantages over existing techniques. In particular, the algorithm is extremely efficient from a computational standpoint, because the initial processing stage requires no convolution-based filtering or template comparison operation. In addition, the algorithm can be implemented to detect and discriminate spikes recorded with signal-to-noise ratios much less than 1.

In the following sections, we briefly review current approaches toward electrophysiological spike discrimination, and present a conceptual overview of phased-array signal analysis techniques within the context of the spike superposition problem. We then derive the equations that form the basis for our phased-array spike sorting algorithm, through which an electrode array can be tuned to respond independently to spikes from different axons in a nerve on the basis of their unique propagation speeds. We evaluate the algorithms using synthetic data that simulate neural recordings from a linear electrode array along a multiaxon nerve, considering several factors that are expected to limit the effectiveness and robustness of the algorithm for separating superimposed spikes. Simulation results demonstrate that this novel phased-array algorithm is very effective and robust in the presence of superimposed spikes and substantial noise.

Current approaches to spike discrimination

A wide variety of approaches for spike discrimination have been described over the last two decades. The most common methods for spike discrimination and identification are based on the analysis of temporal waveforms of the neural spikes (for reviews, see Lewicki 1998; Schmidt 1984). Investigators have used feature analysis (Wheeler and Heetderks 1982), principal-components analysis (Glaser 1968), independent-components analysis (Pham and Cardoso 1999), cluster analysis (Everitt 1997; Zouridakis and Tam 2000), template matching (Salginicoff 1988; Sarna 1988), wavelet transforms (Letelier and Weber 2000; Zouridakis and Tam 2000), signal subspace invariance techniques (Oweiss and Anderson 2002), autoregressive modeling (Liu 1989), neural networks (Chandra and Optican 1997; Jansen 1990), and maximum likelihood approaches (Aitiya 1992; Brillinger 1988). For all of these approaches, the electrode array is treated as a group of spatially independent electrodes, and the part of the information that is related explicitly to the spatial distribution of the spike temporal waveforms with respect to the positions of the electrode contacts in the arrays is not used explicitly. It is important to note that this is also the case for tetrode and “stereo-trode” sorting techniques (Harris et al. 2000; Takahashi et al. 2003), which do not incorporate explicit information about the absolute spacing or positions of the contact points.

A very different approach toward spike discrimination was taken Brown and coworkers (2001). Unlike most earlier investigators, they chose to use the spatial information about composite spike trains recorded with a 2-dimensional photodetector array, rather than using the temporal waveforms of the spikes. By using independent-components analysis (ICA) on the spatial information available from the array, spikes were sorted successfully by resolving the positions in the array from which the signals originated. However, that ICA approach explicitly...
excluded the simultaneous use of information about the temporal structure of the spike waveforms.

A third general approach involves the derivation and implementation of multichannel optimal linear filters (Andreasen et al., 1979; Bierer and Anderson 1999; Gozani and Miller 1994; Roberts and Hartline 1975). That approach incorporates temporal and spatial information about the spikes in the process of spike discrimination. This algorithm is based on the derivation of a set of matched filters for the spike from each different unit on each recording channel. To discriminate 10 units recorded across a 16-channel electrode array, for example, a set of 16 filters would be derived for each of the 10 units to be discriminated, and the entire 16-channel data stream is passed through all 10 sets of 16-channel filters.

A fourth approach, that also uses temporal and spatial information about spikes for discrimination, is based on calculation of the cross-correlation of activity patterns recorded from 2 independent electrodes along a nerve. (Casby et al. 1963; Heetderks and Williams 1975; Williams 1972).

The phased array technique we present here also uses the spatial and temporal information about spike waveforms recorded across various electrode array. As described in more detail in the conclusion, the phased-array algorithm is similar in some respects to the multichannel linear filter approach and to the 2-electrode cross-correlation technique. However, the phased-array approach offers substantial advantages over the standard linear filtering approach, in terms of computational efficiency, robustness to noise, and robustness to spike-waveform nonstationarity.

A new approach: phased-array spike discrimination

Phased-array processing enables the detection, identification, and localization of one or more signals embedded within a background mixture of other signals having similar amplitude and frequency characteristics. Phased arrays have been used in a variety of applications including radar, sonar, communications, imaging, geophysical exploration, astrophysical exploration, and biomedical signal analysis (VanVeen and Buckley 1988). For these applications, a one-dimensional phased array antenna is typically a linear array of N equispaced elements that sample propagating waves in space. The array operates as a spatial filter, selectively receiving a signal from a specific location and attenuating signals from other locations. The algorithm is as follows. Assume that a fixed-frequency incident plane wave projects onto the array from a particular direction of arrival (DOA). The wave crests of the signal will arrive at one end of the array first and will arrive at each successive array element at a progressively later times. The output of the array can be analyzed in such a manner that the signals arriving at the array from that specific DOA can be enhanced, and all those from other DOAs can be attenuated. This is achieved by adding carefully chosen time delays to the signals arriving at the analyzer from the different array elements: if the time delays are chosen to compensate exactly for the time delays of the wave crests from one specific DOA, then summation of all outputs at the analyzer will maximize the component from that DOA. Waves with different DOAs will arrive at each array element with different set of time delays, and summation of the outputs with a set of delays inappropriate for that DOA will result in a much smaller summed signal. In practice, the signal arriving at each sensor is typically multiplied by a complex coefficient to help shape the final pattern produced through summation at the analyzer. The coefficients can be data-independent (as in a classic phased array) or data-dependent (as in an adaptive array).

A single antenna array can be used to differentiate signals from a variety of DOAs simultaneously. This is achieved by configuring the analyzer to carry out simultaneous but independent sets of processing operations, each one corresponding to “tuning” the analyzer to signals from a different DOA. Each of these independent processes is characterized by the addition of a different (and unique) set of phase delays to the signals arriving from the array elements, corresponding to the phase delays of signals arriving from a specific DOA. Similarly, for a radiating phased-array antenna, the majority of the power from the array can be radiated along a particular direction (or set of directions) by controlling the phase of the excitation between the array elements.

We have implemented this approach to enable the discrimination of action potentials propagating along a set of axons in a nerve. We evaluate the simplest case: a one-dimensional electrode array with elements at a uniform spacing along the nerve. These electrode contacts are treated as equivalent to phased-array antenna elements: an action potential traveling through one axon along the nerve will be recorded at successive electrode contacts as a waveform of approximately (but not necessarily exactly) the same shape and amplitude, and will arrive at successive electrode elements at successively increasing time delays. By simply adding the recorded signal streams from all of the electrodes in the array with the corresponding set of propagation delays for that particular axon, the action potentials for that particular neuron will be selectively enhanced over those action potentials propagating at different speeds. Further, because the action potentials propagating through each different axon will have a unique characteristic waveform and propagation speed, a unique set of delays and complex waveform coefficients can be derived for each of the different neural units to be discriminated. Multiple parallel analysis channels can be “tuned” to sort out the spikes from several neurons simultaneously.

In the following sections, we derive the equations that form the basis for our phased array spike sorting algorithm, through which an electrode array can be tuned to respond independently to spikes from different neurons on the basis of their unique propagation velocities. We test the algorithms using synthetic data that simulates neural recordings from a linear electrode array along a multiaxon nerve.

**METHODS**

**Glossary of mathematical symbols**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>d</td>
<td>distance between adjacent electrodes</td>
</tr>
<tr>
<td>e_n(t)</td>
<td>noise in the time domain at electrode n</td>
</tr>
<tr>
<td>F</td>
<td>data sampling frequency</td>
</tr>
<tr>
<td>f_{nm}</td>
<td>the optimal filter derived for the spike of neuron m on channel n</td>
</tr>
<tr>
<td>G</td>
<td>array gain</td>
</tr>
<tr>
<td>j</td>
<td>index for the individual spikes generated by a neuron during the recording window</td>
</tr>
<tr>
<td>K</td>
<td>linear filter order</td>
</tr>
</tbody>
</table>
First, we assume that the structural and biophysical characteristics of the axonal membranes do not vary significantly along the nerve (and electrode array) or with distance along the nerve (and electrode array) per electrode because of conduction delay). Note that the validity of the assumptions to the algorithms. Innovations in methodology may vary significantly with distance along the nerve (and electrode array) or with distance along the nerve (and electrode array) per electrode because of conduction delay). Note that the validity of the assumptions to the algorithms. For the subsequent discussion, we focus on the recorded signal at $x = 0$ for neuron $m$

Consider a linear array of $N$ equispaced recording electrodes, as shown in Fig. 1. Each of the $N$ extracellular electrodes in the array records the spike waveforms generated by $M$ different neurons whose axons project through the nerve beneath that electrode. The task of the algorithm developed here is the decomposition of the mononeuronal composite voltage waveforms recorded with $N$ extracellular electrodes into $M$ sequences of individually resolved action potentials assignable to the $M$ distinct neurons. Spikes in each axon will have a unique propagation velocity $u_i$ and will arrive at the set of electrodes with a unique set of time delays. The $n$th electrode is located at $(n-1)d$, where $d$ is the distance between adjacent electrodes, and receives signal $s_i(t)$

where $e_i(t)$ is the lumped channel noise term. Although the specific values for $u_i$, $v_i(t, x)$ and $e_i(t)$ are not known a priori those values can be obtained from the measured voltages and the recorded data $s_i(t)$. The unknown quantities to be determined through application of the spike discrimination algorithm are the time of occurrence of spikes $\tau_{ij}$

\[ s_i(t) = \sum_{j=1}^{R} \sum_{l=0}^{L} a_{ij} \left[ t - (n-1)d + \tau_{ij} \right] + e_i(t) \quad n = 1, 2, \cdots, N \] (4)

whose $a_{ij}$ are the weights assigned to $x$.

Wave function representation of action potentials

The formalism for our phased-array approach is based on the representation of action potentials as wave functions. To simplify the initial derivation of these wave functions, we start with two assumptions about spike propagation and recording. As discussed in later sections, violations of these assumptions do not invalidate the approach, although some require minor modifications to the algorithms. First, we assume that the structural and biophysical characteristics of the axons and nerve (e.g., axonal diameters, glial sheathing, peak conductances, and activation time constants of the voltage-dependent conductances in the axonal membranes) do not vary significantly over the length of nerve to be analyzed. By "significantly" we mean that the biophysical characteristics are uniform enough to guarantee that 1) the transmembrane potential of a spike in any particular axon will have a nearly invariant waveform in the temporal and spatial domains as it propagates along a nerve, and 2) spikes will propagate along any particular cell’s axon with a constant, characteristic velocity. Second, we assume that the recorded waveform of the spike from any particular axon will appear identical at all electrodes in the recording array (although that waveform will be offset in time at each successive electrode because of conduction delay). Note that the validity of the phased array approach does not depend on rigorous satisfaction of this condition, as will be demonstrated later. Under these conditions, we define the wave function $v_i(t, x)$ for an action potential from an individual neural unit $i$ as a function of time $t$ after the initiation of the spike at the cell’s spike initiation zone (SIZ) and as a function of distance $x$ along the nerve from the SIZ

\[ v_i(t, x) = v_i \left( t - \frac{x}{u_i} \right) \] (1)

where $v$ is the recorded voltage and $u_i$ is the propagation velocity of the spike along cell $i$’s axon. If we substitute a fixed location $x_i$ into Eq. 1, we obtain the temporal waveform at that location. Similarly, if we substitute a fixed time $t_i$ into Eq. 1, we obtain the spatial distribution of the action potential along the axon at that time.

The spike train $V_i(t, x)$ of an individual neuron $m$ during the recording time window is the sum of the propagating wave functions for the individual spikes generated by that cell

\[ V_i(t, x) = \sum_{j=0}^{P_i} v_i(t - \frac{x}{u_i} - \tau_{ij}) \] (2)

where $j$ is the index for the individual spikes generated by a cell $i$, $P_i$ is the total number of spikes generated by neuron $i$ during the recording time window, $\tau_{ij}$ is time of initiation of the $j$th spike from neuron $i$, and $\tau_{0i}$ is initial time delay for the first spike from neuron $i$ to arrive at location $x$.

For the subsequent discussion, we focus on the recorded signal at $x = 0$ for neuron $m$

\[ V_m(t, 0) = \sum_{j=0}^{P_i} s_i(t - \tau_{0i}) \] (3)

FIG. 1. Schematic representation of the phased-array algorithm. A linear array of $N$ extracellular electrodes is placed along a nerve containing several axons. Each of the signal streams $s_i(t)$ is passed through one of $N \times m$ time-delay devices $\{\Delta_{1,1}, \Delta_{2,1}, \cdots, \Delta_{N,m}\}$ used to compensate for the propagation delay of the signals from neuron $m$. Weighting coefficients $\{w_{ij}\}$ are used to shape the output $y_{ij}(t)$. A threshold trigger can be placed after the phased-array output $V_m(t)$ to register the occurrence of a spike.
Phased-array spike discrimination

Following the standard phased-technique, the output of the electrode array can be analyzed in such a manner that all of the spikes arriving at the array from one neuron can be enhanced, and all spikes from other neurons can be attenuated. This is achieved by adding a carefully chosen set of time shifts to the spikes arriving at the analyzer from the different array elements. If the time shifts are chosen to compensate exactly for the time delays of the spikes from one specific neuron, then summation of all outputs will maximize the component from that neuron. Detecting the spike for that unit can then be achieved by simply setting a threshold detector, with a trigger level high enough that only the summation of the appropriately shifted spike records will exceed the threshold. The time of occurrence for that spike can be reported as the time of peak signal amplitude within the portion of the recorded signal above the threshold level.

Figure 1 illustrates this procedure for a single neuron $m$. The time shift $\Delta_{nm}$ added to the signal recorded from electrode $n$ to compensate for the propagation delay for a spike generated by a given neuron $m$ is

$$\Delta_{nm} = \frac{(n-1)d}{u_m} \quad n = 1, 2, \cdots N; m = 1, 2, \cdots M \quad (5)$$

Determination of this set of $N$ time shifts is the fundamental and most crucial step in the practical application of this algorithm, and the approach will not be applicable in situations where a clean set of delays cannot be obtained.

The final estimated waveform for each neuron, $y_m(t)$, is then the average of the time-shifted waveforms recorded from all electrodes

$$y_m(t) = \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{M} \sum_{j=0}^{r_i} \left[ t - \left( \frac{(n-1)d}{u_i} - t_r - \frac{(n-1)d}{u_n} \right) \right]$$

where the set of $N$ values of $w_{nm}$ for each unit $m$ are scalar weighting factors to equalize the effective gains across all channels. This ensures equal weighting of the signals from all $N$ channels in the final shifted and summed output for that unit. Note that, in practice, the relative amplitudes of different units on different channels of nerve recordings is often invariant, and that the assignment of appropriate uniform weighting factors for all units can be accomplished simply by adjusting the recording gains on the different channels. For simplicity in the subsequent derivations, we assume that the relative unit amplitudes are uniform across channels, and all channel gains are approximately equal. In this case, these weight coefficients reduce to $1/N$ for all channels. Equation 6 then becomes

$$y_m(t) = \frac{1}{N} \sum_{n=1}^{N} u_n(t + \Delta_{nm}) \quad m = 1, 2 \cdots M \quad (6)$$

Substituting Eq. 4 and Eq. 5 into Eq. 7, we obtain

$$y_m(t) = \frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{r_i} \sum_{j=0}^{r_j} \left[ t - \frac{(n-1)d}{u_i} - t_r - \frac{(n-1)d}{u_n} \right]$$

$$+ \left[ t - \frac{(n-1)d}{u_i} \right]$$

$$+ e_i \left[ t - \frac{(n-1)d}{u_i} \right]$$

By separating the right side of Eq. 8 into 2 parts (i.e., the first part for $i = m$ and the second part for $i \neq m$), we obtain a new formulation for the application of the phased array method to neural data

$$y_m(t) = \sum_{j=0}^{r_i} u_n(t - t_r) + \frac{1}{N} \sum_{n=1}^{N} \sum_{j=0}^{r_i} \sum_{j=0}^{r_j} \left[ t - \frac{(n-1)d}{u_i} + \frac{(n-1)d}{u_n} - t_r \right]$$

$$+ \frac{1}{N} \sum_{n=1}^{N} \sum_{j=0}^{r_i} \left[ t - \frac{(n-1)d}{u_i} \right]$$

The first term in Eq. 9 corresponds to the average of the optimally shifted (i.e., temporally superimposed) spike waveforms from neuron $m$, referenced to location $x = 0$ (i.e., the first electrode of the recording array). Note that the amplitude and time course of this signal component will be approximately equal to the spike signal recorded on each of the individual electrodes; that is, the spike signals to which this analyzer has been “tuned” will be neither temporally dispersed nor divisively attenuated by this averaging operation.

The second term in Eq. 9 corresponds to the interference of the signals from other neurons recorded by all electrodes. These spike signals will not have been superimposed precisely. Therefore, their signal component resulting from this averaging operation will be dispersed in time, and will be lower in amplitude than the corresponding signals at each of the individual electrodes by a factor of up to $1/N$.

The third term in Eq. 9 corresponds to the average of the noise recorded at all electrodes. Note that if this recording noise is truly independent at each of the $N$ electrodes, it will be reduced through this averaging operation by a factor of approximately $1/ \sqrt{N}$. Thus, the analyzed signal will have substantially improved signal-to-noise level with respect to the signals recorded at the individual electrodes.

Performing the discrimination operation

The ability of the algorithm to discriminate the in-phase signals from the background signals in the presence of noise depends on the relative amplitudes of these 3 different signal components. As we will show below, it is possible to configure a recording array such that the amplitude of the first component (i.e., the average of the optimally shifted spike waveforms from the “target” unit corresponding to the first term of Eq. 9) will reliably exceed the amplitude of the second term (even in the presence of multiple superpositions and considerable noise, represented by the third term), if the number of electrodes in the array can be increased indefinitely and if an adequate length of nerve is accessible. Given an adequate number of electrodes, the discrimination operation can be reduced to the task of defining a simple threshold detector, with a trigger level that is greater than the peak amplitude of the second-term component of Eq. 9 and less than the amplitude of the first-term component.

Note that this algorithm can be applied to any subset (or all) of the $M$ units recorded on the set of $N$ electrodes simultaneously, given adequate signal-to-noise characteristics for the individual units. This is achieved by deriving $M$ sets of $N$ time shifts corresponding to each of the $M$ different sets of spikes from the $M$ axons recorded by the array of $N$ electrodes. Simultaneous, independent computational processes can then be initiated to operate on the same input data. Further, it should be noted that this approach could be implemented to operate online, in real time, with appropriate hardware.

To illustrate one substantial problem that constrains the effectiveness of this (and every other) spike discrimination approach, consider a case in which the experimenter wishes to discriminate a set of spikes with amplitudes that span an order of magnitude. How effective will our approach be for such cases? In other words, can we configure an electrode array and phased-array analyzers such that the analyzed signal attributed to the first term in Eq. 9 will be larger than the signal attributed to the second term, in the presence of realistic noise levels, even in cases where the spike amplitude of a background unit might be 10 times larger than the spike from the target unit? From the second and third terms of Eq. 9 we see that the background signal interference will decrease $\approx 1/N$ and the effective noise power for sorted spikes will decrease $\approx 1/ \sqrt{N}$. Thus, if the peak amplitude of one

\footnote{In our experience, this particular task is extremely problematic for spike discriminators based on multichannel linear filters. The passage of a large-amplitude spike waveform through a filter that was tuned to a spike with much smaller amplitude can result in a spurious transient response that is as large as (or even larger than) that filter’s response to the “correct” (matched) unit. Such spurious transients are reported as false-positives.}
Determining the minimal number of electrodes needed for effective discrimination

In general, the discrimination effectiveness using this algorithm will improve with the addition of more electrodes. However, technical considerations and/or the size scale of the biological preparation may constrain the array size, favoring an array with the minimal number of electrodes to achieve a particular operational criterion. One practical approach toward determining the minimal number of electrodes needed to achieve robust discrimination is presented in the following sections.

Consider the composite spike and noise signals arriving at any single electrode in a recording array. The arrival of one or more spikes from neurons other than \( m \) at that single electrode would create a transient signal that would ultimately contribute to a composite background signal at the analyzer for cell \( m \). We wish to determine the number of such electrodes we will need in the array to implement a simple threshold detector for discriminating a spike from cell \( m \) above the composite background signal calculated at the analyzer. That is, how many channels would we need to average (with appropriate delays) to guarantee that the smallest unit we wish to discriminate will sum to a larger signal than the background transient from the largest unit present in the recordings? We define the amplitude of largest recorded signal on any single channel to be

\[
|s|_{\text{max}} = \max \left\{ \sum_{i=1}^{n} \sum_{j=0}^{r} \left[ i \cdot \left( \frac{n-1}{n} \cdot \frac{1}{u_i} - \tau_j \right) \right] + \epsilon(t), \quad n = 1, 2, \ldots, N \right\}
\]

(10)

We can then define an effective detection threshold \( \alpha \) as

\[
1 \geq \alpha = \frac{|s|_{\text{max}}}{N \cdot |V_m|}
\]

(11)

where \( |V_m| \) is the amplitude of the spike from neuron \( m \) on that same channel. That is, the experimenter would need to implement a large enough array \( N \) to guarantee that \( |s|_{\text{max}} \) was significantly less than \( N \cdot |V_m| \), and would then select a value for \( \alpha \) that was intermediate between \( |s|_{\text{max}} \) and \( N \cdot |V_m| \). In practice, the values for \( |s|_{\text{max}} \) and \( |V_m| \) would be determined from an examination of the recorded data, and would depend on the range of individual unit amplitudes and the frequency with which multispike superposition occurred. Given values for \( |s|_{\text{max}} \) and \( |V_m| \), a value for \( N \) can be selected to obtain a specified value for \( \alpha \)

\[
N \geq \frac{|s|_{\text{max}}}{\alpha|V_m|}
\]

(12)

The selection of a robust value for \( \alpha \) would also be determined operationally: for example, the experimenter might determine that robust discrimination ultimately requires 1) a typical difference of 33% in amplitude between the smallest unit’s summed response and the largest unit’s spurious background transient, and 2) a threshold of \( \alpha = 0.75 \), that is, so that all transients passing through the analyzer having an amplitude greater than 0.75\( |V_m| \) would be reported as a spike from neuron \( m \), whereas all signals with amplitude less than or equal to 0.75\( |V_m| \) would be treated as background signal. In the scenario described earlier, where the amplitude of recorded signal at a channel might be 10 times the amplitude of unit \( m \), then the experimenter would require an electrode array with at least

\[
N \geq \frac{|s|_{\text{max}}(|V_m| \cdot \alpha)}{10/0.75} = 13.3.
\]

Thus, a minimum of 14 electrodes would be required.

ADDITIONAL CONSTRAINTS IMPOSED BY RECORDING NOISE. If there is significant noise in the recorded signals, the contribution of the third term in Eq. 9 must also be considered in the determination of the number of electrodes: more recording points than the minimal number calculated above might be needed to average out the noise to achieve adequate \( \text{SNR} \). We start the consideration of array signal to noise ratio \( \text{SNR}_{\text{array}} \), by defining the array gain \( G \) as the ratio of the \( \text{SNR}_{\text{array}} \) to the sensor signal to noise ratio \( \text{SNR}_{\text{sens}} \) (Johnson and Dudgeon 1993)

\[
G = \frac{\text{SNR}_{\text{array}}}{\text{SNR}_{\text{sens}}}
\]

(13)

If the channel gains are all approximately equal, as we assumed for our derivation, then these weight coefficients reduce to 1/\( N \) for all channels, the gain achievable through phased-array processing can then be written as

\[
G = \frac{\sum_{i=1}^{N} |V_i|^2}{\sum_{i=1}^{N} |V_i|^2} = N
\]

(14)

From Eq. 13 and Eq. 14, we obtain

\[
\text{SNR}_{\text{array}} = N \cdot \text{SNR}_{\text{sens}}
\]

(15)

Thus, an intrinsic feature of phased array processing is the improvement of \( \text{SNR} \) at the output of the analyzer with respect to the \( \text{SNR} \) at the sensor by a factor of \( N \).

To consider the effect of \( \text{SNR} \) on determining the required number of electrodes in the array, we define a threshold criterion for the \( \text{SNR} \) we wish to achieve as \( \text{SNR}_{\text{array}} \geq \text{SNR}_{\text{required}} \). Using Eq. 15, this additional criterion can be rewritten as the following constraint on the number of channels

\[
N \geq \frac{\text{SNR}_{\text{required}}}{\text{SNR}_{\text{sens}}}
\]

(16)

Thus Eq. 16 and Eq. 12 represent the 2 principal constraints on the number of recording channels that must be met for effective operation of the phased-array technique for a unit having a particular \( \text{SNR}_{\text{sens}} \) and amplitude ratio to the largest unit, respectively.

To demonstrate the application of this additional noise-related constraint, consider the effect of adding noise to the recordings used in the preceding example, where the amplitude of recorded signal at a channel is 10 times the amplitude of unit \( m \). Specifically, we will assume that the recording traces have an observed \( \text{SNR}_{\text{sens}} \) of 1 (0 dB) for the largest unit, which would mean that the smallest unit would be buried in the noise with an \( \text{SNR}_{\text{sens}} \) of only 0.1 (~10 dB). Assume further that the operator wishes to achieve a minimum \( \text{SNR}_{\text{required}} \) of 1 (0 dB) for the smallest unit. In this case, Eq. 12 yields \( N = 14 \) and Eq. 16 yields \( N \geq 10/1 = 10 \). Thus, even in the presence of considerable noise, the constraint imposed by Eq. 12 dominates the determination of electrode number: only 14 electrodes would be needed to satisfy that constraint, and the noise reduction criterion would automatically be satisfied. To extend this illustration, consider this case if \( \text{SNR}_{\text{required}} \) was 2 (3 dB). In this case, solving Eq. 16 would yield a value of 20 electrodes as the minimum required number in the array.
Discriminating superimposed spikes having similar conduction velocities

The phased-array algorithm discriminates spikes based largely on differences in their conduction velocities. If spikes from 2 different axons had precisely the same conduction velocities, then this algorithm would fail to discriminate instances of their superposition (unless additional filtering operations were added to the procedure to discriminate the superimposed spikes on the basis of differences in their waveforms). What is the minimal difference between the conduction velocities of spikes in 2 different axons that would be necessary for the phased-array algorithm to resolve superposition?

Consider 2 spikes propagating along 2 different axons j and k in a nerve, recorded with a linear electrode array having N electrodes distributed along a total length L. Assume that the spikes have conduction speeds $u_j$ and $u_k$, where $u_j > u_k$, and $\Delta u_{jk} = |u_j - u_k| \ll u_k, u_j$. The propagation times for the spikes to travel along the entire array are defined as $T_j$ and $T_k$, respectively. The temporal half-widths of the spikes are defined as $(\tau_{1/2})_j$ and $(\tau_{1/2})_k$. For simplicity, we will assume that the temporal half-widths of spikes having similar conduction speeds are equal, and therefore we assume that $(\tau_{1/2})_j = (\tau_{1/2})_k = (\tau_{1/2}).$ The phased-array algorithm will allow discrimination of the 2 spike waveforms if the distance between their peaks is larger than their temporal half-widths, after propagating half of the length of the electrode array (i.e., we assume the possibility of occurrence of the “worst case” superposition scenario, where the 2 spike waveforms are superimposed perfectly at the center electrode in the array). To meet this constraint, the spikes would have to satisfy the following condition

$$\Delta u_{jk} = |u_j - u_k| \ll u_k, u_j.$$  \hspace{1cm} (17)

Equation 17 can be further reduced to obtain a simple expression for the resolution of the algorithm with respect to the minimal fractional difference in spike conduction velocities that can be discriminated

$$\frac{\Delta u_{jk}}{u_{jk}} \geq \frac{2\tau_{1/2}}{L} \quad \text{or} \quad u_{jk} \leq \frac{2\tau_{1/2}}{L}.$$  \hspace{1cm} (18)

To invert the problem, we can also use this formula to calculate the value L that would be required for an electrode array capable of discriminating superimposed spikes having a specific set of conduction velocities.

RESULTS

We evaluated the performance of the phased-array spike discrimination algorithm using synthetic data that simulated neural recordings from a linear electrode array along a multi-axon nerve. The simulated spike trains were based on multichannel electrophysiological recordings of the abdominal nerve cord of the cricket *Acheta domestica* (Spence and Hoy 2003). In that study, the recording array consisted of an equispaced set of 16 silicon-based electrodes, with interelectrode spacing of $d = 600$ μm. The nerve cord contains hundreds of axons. However, all but approximately 20 of those axons are so small that their spikes cannot be discerned with whole nerve extracellular electrodes. The spikes from these 20 large axons are easily discernible, with SNR as high as 100 using conventional amplifiers. These axons are from the “giant” mechanosensory interneurons of the cercal sensory system, which respond to air current stimuli.

Based on isolated (i.e., nonoverlapping) instances of spike waveforms recorded by Spence and Hoy from 4 of these giant axons during those experiments, we constructed a set of 16-channel spike waveform “templates” for each of these 4 different units. Each of the 4 16-channel templates corresponded to the voltage traces that would have been recorded across the array of 16 electrodes as the simulated spike propagated along the nerve. The 4 units had conduction velocities of 5, 4, 3, and 2 m/s, which bracketed the entire range observed experimentally. The relative amplitudes of the four units were 1.0, 0.8, 0.6, and 0.4, respectively. In all figures, these units will be identified with numerical labels 1 through 4, starting with the neuron having the fastest (and largest amplitude) spike, and progressing to the neuron with the slowest (and smallest amplitude) spike.

Figure 2A shows the simulated signals for all 4 of these units, as they would appear at 4 of the 16 electrodes in the array (labeled as Ch 1, Ch 4, Ch 8, and Ch 13). In this and all subsequent simulations, we generated the synthetic multunit spike trains by summing the 16-channel templates for the 4 different units, each offset from the others by a chosen time interval. (For some later simulations, simulated channel noise was also added to the traces.) In this figure panel, the 4 units occur in descending order relative to their amplitude, and none of the spikes was superimposed with any other spikes at any of the electrodes.

**Demonstration of the phased-array spike-sorting technique**

Figure 2B shows the result of passing the 16 channels of simulated data through the set of 4 phased-array analyzers for the 4 different units. The top 4 traces are the processed signals for neuron 1, neuron 2, neuron 3, and neuron 4, labeled $y_1(k)$, $y_2(k)$, $y_3(k)$, and $y_4(k)$, respectively. The bottom trace is the reference composite signal at channel 1 (i.e., the simulated
multunit recording at electrode 1). Note that the analyzer trace for each unit has a strong transient at the time of occurrence of its corresponding template in the composite waveform, and that the resulting transient looks identical in shape to the corresponding simulated spike on recording channel 1. This is precisely what is expected from the phased-array algorithm: each analyzer’s response is obtained by 1) shifting the traces recorded on each of the 16 electrodes by amounts equal to the cumulative conduction time between each electrode and all successive electrodes up to \( n = 16 \), 2) summing all appropriately shifted signals, and 3) dividing the sum by 16. Thus, if the shape and amplitude of the unit is the same on all 16 channels, that original shape and amplitude will be recovered through the analysis. In practice, the shapes and amplitudes might differ slightly between electrodes, so that the final transient produced at the analyzer would actually be the average of all 16 recorded spike shapes. Note also that each analyzer trace has an extended “ripple” spread out around the times of occurrence of the other 3 units. This “interference” ripple is also expected: this ripple is the result of summing 16 inappropriately shifted instances of each background unit, and dividing those nonoverlapping signals by 16. This is most easily seen on the analyzer for unit 4: the first set of ripples is attributed to the nonoverlapping summation (and division by 16) of the largest unit (1) as it passed through the analyzer tuned to the propagation speed for unit 4.

It is clear from this illustration that each of the units would be correctly classified by a simple thresholding operation on the analyzer trace. Even for the smallest unit (4), the amplitudes of the transient ripples corresponding to the largest unit (1) were lower in amplitude than the single transient caused by unit 4.

**Discrimination for two spikes with similar velocities and waveforms**

We tested the limit of the algorithm’s capability to discriminate 2 spikes having identical half-widths and very similar conduction velocities. To do so, we selected unit 1 (i.e., the fastest spike) as the test case, and derived the conduction velocity for a spike that would be right at the limit of discriminability from unit 1, according to Eq. 18. Unit 1 had a spike with half-width \( r_{1/2} \) = 100 \( \mu \)s and speed \( u = 5 \) m/s. Substituting these values into Eq. 18, along with the standard values of \( N = 16 \) and \( d = 600 \) \( \mu \)m for the electrode array, we calculate the velocity of the next-fastest spike that could be discriminated from unit 1 would be 4.89 m/s.

Figure 3A presents the results of a simulation in which the phased-array algorithm was applied to these 2 units. \( V_1(k, 0) \) is a simulated segment of the output from unit 1, with a conduction speed \( u_1 = 5 \) m/s. This segment of the simulated recording contained 2 spikes. \( V_2(k, 0) \) is a simulated segment of the output from a second test unit, with a conduction speed \( u_2 \) of 4 m/s. That cell also fired 2 spikes during the simulated segment. The first spikes from the 2 different cells were precisely superimposed at the first electrode, as shown in the third trace, which represents the multunit composite recording of both units as they would appear on channel 1. The second spike from each cell occurred later in the window, and neither of these second spikes overlapped with a spike from the other cell. Fourth trace, labeled \( y_j(k) \), shows the output of the analyzer for the top template trace \( V_1(k) \), and the bottom trace labeled \( y_k(k) \) shows the output of the analyzer for the second template trace \( V_2(k, 0) \). Both units could be discriminated correctly with a thresholding operation.

**Discrimination of superimposed spikes**

A major motivation for the development of this algorithm was to enable the detection, discrimination, and identification of multiply superimposed spikes from multichannel extracellular recordings.
Innovative Methodology

FIG. 4. Application of the phased-array algorithm to a simulated 16-channel recording containing superimposed spikes from 4 units. A: colored traces in the left panels are the individual templates of the 4 different units that were summed to construct the simulated 4-unit records, which are shown as the black traces in the right set of panels. Records from only 4 of the 16 channels are shown (i.e., channels 1, 4, 8, and 13.) Timing of the spikes was adjusted to coincide in their arrival times at the center electrode (i.e., 8) of the recording array. B: results of passing all 16 channels of simulated data through each of the 4 analyzers. Top 4 lines are the outputs of the 4 analyzers, color-coded as in the left panels in A. For reference, the composite waveform simulated at channel 1 is plotted on the bottom trace.

Lar nerve recordings. The simulations in this section test and illustrate the robustness of this algorithm under conditions of multiple spike superposition. Figure 4 illustrates the results of the application of the algorithm to a simulated 16-channel recording of superimposed spikes from all 4 units. The conduction speeds were $u_1 = 5$ m/s, $u_2 = 4$ m/s, $u_3 = 3$ m/s, and $u_4 = 2$ m/s. The colored traces in the left panels of Fig. 4A are the individual templates of the 4 different units that were summed to construct the simulated 4-unit records, which are shown as the black traces in the right set of panels in Fig. 4A. Note that we show records from only 4 of the 16 channels (i.e., channels 1, 4, 8, and 13.) For this simulation, the timing of the spikes was adjusted to the “worst case” superposition scenario: the case for which the spikes coincided in their arrival times at the center electrode (channel 8) of the recording array. This resulted in the least possible degree of temporal separation at either end of the recording array.

The phased-array algorithm was then applied to the simulated recordings, to evaluate its effectiveness in discriminating the superimposed spikes. Figure 4B shows the results for each of the 4 analyzers. The top 4 lines are the outputs of the 4 analyzers, color-coded as in Fig. 4A. For reference, the composite waveform simulated at channel 1 is plotted on the bottom trace. For this illustration, the algorithm had been configured to report the spike times referred to the recordings on that first electrode, which is the site closest to the spike initiation zone of the axons. We note that the algorithm could also be configured to report the spike timing closest to the postsynaptic target end of the electrode array. It is clear from this illustration that the identity and time of occurrence of all 4 of the units would be reported reliably through a simple thresholding operation (and subsequent peak-locating operation) on the output of the analyzers.

Illustration of the robustness of the phased array algorithm to superposition of large units with a small unit

In our experience, a significant limitation to the robustness of spike discrimination based on multichannel linear filtering approaches lies in the generation of false-positive identification of small units, when spikes from a large unit are passed through the filter for those small units (Gozani and Miller 1994). Even with an optimally-configured matched multichannel filter for a small unit, passage of a very large amplitude “unmatched” spike through the filter set matched to the small unit will yield a transient of sufficient amplitude to register a false positive for the small unit. The simulations in this section test the robustness of this phased-array algorithm under these conditions.

Figure 5 illustrates the results of the application of the algorithm to a simulated 16-channel recording. As in Fig. 4, results are shown for only 4 of the 16 channels. The colored traces in Fig. 5A are the individual templates of 4 different units that were summed to construct the composite multunit records, which are shown as the black traces in Fig. 5B. Note that for this simulation only, one of the templates (identified as unit 1) was set to have an amplitude that was twice the amplitude of the largest unit seen in the studies by Spence and Hoy. This was done to exaggerate the possible artifact of passing a large unit through an analyzer for a small unit. The other units (i.e., numbers 2, 3, and 4) were the same as for the preceding (and subsequent) simulations. In this simulation, each of the 3 largest units were set to fire 2 spikes. The first spike from each of these 3 cells did not overlap with any of the others, but the second spikes from all of these 3 cells were set to coincide in their arrival at the center electrode in the array. The smallest unit was set to fire only one spike, near the center of the time segment.

Figure 5C shows the results of applying the phased-array algorithm to these data. The top 4 colored traces are the analyzer outputs for units 1, 2, 3, and 4, respectively. The bottom (black) trace is the simulated composite recording at electrode 1 for reference. The fourth trace [labeled $y_4(t)$] is the analyzer for the smallest unit. Although the other units cause the expected spurious “ripple” transients on this analyzer, none of those other units (nor their superposition) results in transients that are as large as the correct unit, and thus a simple threshold discriminator could be set to operate effectively.

Illustration of the robustness of the algorithms to noise

As discussed in Methods, an essential stage of the phased-array processing of N channels of data is the averaging of all N time-shifted records. Thus, if the recording noise is truly independent at each of the N electrodes, the noise at each
The analyzer’s output will be reduced by a factor of approximately $1/\sqrt{N}$ with respect to the noise on the individual input channels. Figure 6 illustrates the robustness of the phased-array algorithm to the addition of noise. For this simulation, we repeated the simulation that formed the basis for Fig. 4, but added zero-mean Gaussian white noise to all recording channels, with an amplitude sufficient to reduce the signal to noise ratio (SNR$_{sensor}$) to 10 dB at each channel. The noise was independent at all channels. The 4 traces in Fig. 6A show simulated recordings at channels 1, 4, 8, and 13 out of the 16 total channels in the array, during a segment in which all 4 cells fired a spike. As in the right set of panels of Fig. 4, the units were set to coincide precisely in time at electrode 8. In these noisy simulated recordings, the largest unit is clearly discernable at all 4 electrodes. However, the smaller units are nearly indistinguishable from the noise, and would certainly be difficult to discern above noise as a candidate spike, with an automated event detector of the type usually implemented in commonly available spike-sorting software packages.

Figure 6B shows the results of passing these noisy 16-channel recordings through the phased-array analyzers for the 4 units, and can be compared directly with Fig. 4B. The top 4 lines are the outputs of the 4 analyzers, color-coded as in Fig. 4A. For reference, the composite waveform simulated at channel 8 is shown in the bottom trace.
illustration of the detection of units with SNR_{sensor} < 1

When the SNR_{sensor} for a particular unit is below one, that unit is effectively invisible to all spike discrimination protocols that rely on thresholding for the preliminary identification of candidate spikes. However, because the summation of N time-shifted signals suppresses uncorrelated noise by an order of 1/√N, the phased-array algorithm can actually enable the detection of units that have amplitudes with SNR_{sensor} well below unity. This represents a major, fundamental advantage of this phased-array approach over other methods.

To achieve this capability, the phased-array algorithm needs to be amended with a protocol for “scanning” through the multichannel data records with candidate sets of delays, and identifying those sets of delays that bring out units buried in the noisy recordings. A simple “blind event detection” algorithm would be as follows.

1) A candidate set of delays is programmed into a phased-array analyzer.

2) The multichannel data is passed through the analyzer, and the resultant output signal is processed with a threshold detector or other reasonable event identifier.

3) If a significant event is not detected, then the set of delays is incremented by a very small value, and the process loops back to step 1 above.

4) If a significant event is detected, then that set of delays is registered as corresponding to a unit.

5) The process loops back to step 1 with a minimally incremented set of delays, to search for additional units.

Figure 7 presents an illustrative example. Assume that SNR_{sensor} for one small unit across our standard 16-element linear array is 0.1, or −10 dB. The 4 red traces in Fig. 7A are noise-free waveforms for that unit at 4 of the 16 channels, and the black traces are the results of adding Gaussian white noise to those waveforms to obtain the target SNR. Figure 7B shows the result of passing the 16 noisy channels through the phased-array analyzer set with the appropriate delays for that unit. The top trace is the phased-array–processed data referenced to channel 1, and the second trace is the template without added noise for channel 1 for comparison. It is clear that this algorithm would enable detection of this unit.

Determining the minimal number of electrodes needed for effective discrimination

Figure 7 demonstrated that a unit with SNR_{sensor} of 0.1 dB could be discriminated with the standard array of 16 electrodes. How would the effectiveness of the algorithm be influenced by changing the number of recording sites? In the section leading up to Eq. 16, we derived a simple metric for setting the minimum number of array elements required for reliable discrimination, under several relevant constraints, including the SNR_{sensor} of the smallest unit and the ratio in amplitudes of the smallest and largest units to be discriminated. Here we apply those constraints to the set of 4 test units used for Fig. 4, under

FIG. 6. Application of the phased-array algorithm to a simulated noisy 16-channel recording containing superimposed spikes from 4 units. A: templates of the 4 different units were summed and added to zero-mean Gaussian white noise to construct the simulated 4-unit records. Smaller units are nearly indistinguishable from the noise. Records from only 4 of the 16 channels are shown (i.e., channels 1, 4, 8, and 13.) As in Fig. 4, the timing of the spikes was adjusted to coincide at the center electrode (i.e., 8) of the recording array. B: results of passing all 16 channels of simulated data through each of the 4 analyzers. Top 4 lines are the outputs of the 4 analyzers, color-coded as in the left panels in Fig. 4A. For reference, the composite waveform without added noise, simulated at channel 1, is plotted on the bottom trace.

FIG. 7. Application of the phased-array algorithm to discover a unit with amplitude below the recording noise in a simulated 16-channel recording. A: A 4 red traces are noise-free waveforms for the small unit at 4 of the 16 channels, and the black traces are the results of adding Gaussian white noise to those waveforms. It is clear that this unit is buried in the noise at all electrodes, and would not be detectable with standard transient event detectors. B: result of passing the 16 noisy channels through the phased-array analyzer set with the appropriate delays for the small unit. Top trace is the phased-array–processed data referenced to channel 1, and the second trace is the template plus noise for channel 1 for comparison.
the noise conditions illustrated in Fig. 7A [i.e., the SNR$_{\text{required}}$ for the smallest unit is 0.1 − 10 dB). We impose the constraints that $\alpha = 0.7$ and SNR$_{\text{required}} = 1$ (0 dB) for the smallest unit. The relative amplitudes of the 4 units were 1.0, 0.8, 0.6, and 0.4, respectively, as in the previous simulation. We simulate a “worst case” superposition, in which all units are precisely superimposed at the center electrode in the array. This superposition results in a composite waveform with peak amplitude which is 7 times larger than the amplitude of the smallest unit. Using Eq. 12 and Eq. 16, the minimal number of channels is determined to be 10.

Figure 8 illustrates the results of the analysis using 3 different simulated recording arrays, having 8, 16, and 32 electrodes at interelectrode spacings of 1200, 600, and 300 μm, respectively. The first trace $[y_n(t), N = 8]$ corresponds to the simulated array with only 10 sensors. This is less than the critical number needed for discrimination according to our criteria, and the smallest unit cannot be discriminated above the noisy transients associated with unit 1. The second trace $[y_n(t), N = 16]$ corresponds to the simulated array with the critical number of sensors. The smallest unit can just be discriminated above the noisy transients associated with unit 1. The third trace $[y_n(t), N = 32]$ shows the results using double the required number of channels. In that case, the smallest unit is clearly discriminable. The bottom trace is the template for the smallest spikes without added noise for channel 1 for comparison.

DISCUSSION

The phased-array approach described here was developed to detect and classify spikes from multiple, simultaneously active neurons from multichannel recordings, even in situations where there was a high degree of spike waveform superposition and channel noise. The analysis of simulated composite spike trains presented here demonstrates that the algorithm is, in fact, capable of reliably discriminating spike trains under conditions expected during actual physiological recordings. It is important to note that the technique is applicable only to situations in which each different unit to be discriminated is recorded at 2 or more recording sites, where some degree of propagation delay exists between the recording sites. For example, multiple channel recordings of a single unit from a single tetrode will not suffice. The protocols are best suited to nerve preparations, in which several recording points can be established along a length of nerve containing the axons of all cells to be discriminated. We note further that the algorithm’s ability to resolve superimposed spikes depends on the number of independent electrodes with which spikes from any particular axon can be recorded. Although all derivations and illustrations were done within the context of conventional electrophysiological electrode arrays, the protocols may be well suited to optical recording techniques, where the number and arrangement of the effective recording points is not limited by the physical dimensions of an array of electrodes. Finally, it is important to note that the application of this algorithm requires an accurate determination of the set of interelectrode propagation delay times for each unit to be discriminated. The inability to determine a unit’s propagation speed would defeat the algorithm’s ability to detect that unit, and could confound the discrimination of other units having similar speeds. In practice, many clean examples of the spike from a neuron might be required to derive a robust set of delays. This would be difficult in situations where there is a high level of baseline activity, although the design of training stimuli crafted to minimize spike superpositions might help to mitigate this problem.

In the following sections we consider the assumptions on which our derivations were based, we contrast the phased-array approach to another spike discrimination approach that also uses the spatial and temporal information about action potential waveforms, and we discuss the relative advantages and limitations of the phased-array approach.

Consideration of underlying assumptions

Several assumptions were made to simplify our derivation of the equations for this phased-array approach, and the constraints imposed by those assumptions would be expected to present practical limitations to the applicability of the algorithm.

UNIFORMITY OF SPIKE SHAPE AND AMPLITUDE ACROSS ELECTRODES. First, we assumed that the spike amplitude and waveform from any single unit was identical at all electrodes across the array. If that condition is satisfied, then the waveform at the output of each analyzer is identical to the spike waveform. However, this is not an essential condition for effective operation of the algorithm. In actual experimental conditions, there might be significant differences in the shape and amplitude of the recorded waveforms for a specific unit at different electrodes, attributed to differences in the electrode/nerve interface or to real differences in axon diameter at different locations along a nerve. In this case, the waveform at the output of the phased-array analyzer for each unit would be the average of the shapes at all electrodes. Such variation of the shape across electrodes would not necessarily degrade performance. We expect that the effectiveness of the algorithm could be optimized in such circumstances by scaling the different gains on all recording channels to a common value before the first stage of phased-
array analysis, to achieve uniform weighting across all of the available channels.

**Uniformity of Propagation Delay Between Electrodes.** A second assumption made during our derivation was that the propagation delay for each cell’s spike between each successive pair of recording points was identical. However, several factors might lead to a variation in propagation delays between successive electrodes. The factors might be biological in origin (e.g., a significant variation in axon diameter or glial sheathing along a nerve) or technical (e.g., nonuniform stretching of the preparation during mounting of the nerve along the electrode array). Here again, this assumption of uniform delay is not critical for optimal effectiveness of the algorithm. The essential requirement for the algorithm is that the unique set of interelectrode propagation delays be determined for each unit, and programmed into the analysis protocols.

Operationally, this could be a very straightforward task, and could be incorporated into the unit identification stage of any practical implementation of the algorithm, presented in association with Fig. 7. Specifically, referring to the protocol for “blind scanning” for units with low SNR, the following processing step could be added after step 3:

4’ If a significant event is detected for a specific set of test delays, then each of the individual interelectrode delay values is “jittered” around the initial (uniform) value, and the amplitude of the final output of the analyzer is assessed. The set of delay values that gives the maximum analyzer output is registered as the optimal operational set of delays for that unit. Note that the ultimate operational speed of this algorithm does not depend on the set of delays being uniform across electrodes: the efficiency of the computations involved in applying a set of delay lines is not affected by the durations of those delays.

**Stationarity of Spike Waveform and Propagation Speed.** A third assumption made during our derivation was that the waveform and propagation speed of each neuron’s action potential are stationary during the experiment. Any significant variation of either parameter would be expected to compromise the effectiveness of this (or any other) spike discrimination procedure.

Physiologically relevant situations are known to exist in which spike amplitudes are known to vary over time. First, variation in the quality of the electrode-to-nerve interface during the course of an experiment could result in significant changes in apparent spike shape. Second, during maintained, high-frequency bursts of activity, the action potentials generated later in the burst can in some cases have significantly lower amplitude than spikes occurring at the beginning of the burst. Third, various types of interactions between fibers in bundles of unmyelinated axons have been reported in some preparations (Andersen and Yang 1989). These interactions might occur when the membrane potential of one fiber is significantly influenced by activity in nearby axons ascribed to their extracellular fields. However, as summarized above, because the effectiveness of the phased-array algorithm does not depend on the precise shape of spikes, nonstationarity of spike waveform should not present a significant problem for this approach. This is, in fact, another significant practical advantage of the phased array technique over conventional linear filtering or template-matching approaches.

Nonstationarity in conduction velocity is, however, problematic for the phased-array approach. It would be essential to detect any drift in spike conduction speed and to correct for shifts in interelectrode delay times through implementation of some type of adaptive algorithm.

**Comparison of the Phased-Array Approach to Multichannel Matched Linear Filtering**

As discussed in the introduction, the phased-array algorithm uses spatial and temporal information about the spike waveforms for discrimination, and thus has advantages over algorithms limited to either the temporal or spatial information alone. Another common approach also incorporates temporal and spatial information about spikes in the process of spike discrimination. That algorithm is based on the derivation of a set of matched, multichannel filters for the spike from each different unit on each recording channel. The specific implementation of the phased-array approach enables several substantial computational advantages over the conventional filtering approach, at the expense of a reduction in sensitivity to differences in spike waveform that can aid discrimination between spikes having similar conduction speeds. An understanding of the relationship of these two approaches sheds light on their relative merits.

Discrimination of a spike based on optimal matched filtering is achieved by computing the correlation between a filter for that spike and the neurophysiological data (Oğuztöreli and Stein 1977)

$$\tilde{y}_m(t) = \sum_{n=1}^{N} f_{mn}(t)s_n(t)$$

(19)

where $f_{mn}(t)$ is an optimal filter derived for the spike of neuron $m$ on channel $n$. This correlation operation is the most computationally intensive step in the procedure. The resulting outputs $\tilde{y}_m(t)$ of the filtering operation are then typically passed through a threshold detector, and those filter transients that surpass the threshold are registered as a spike from neuron $m$.

This functional representation of the linear filtering approach can be compared with the equivalent functional representation of the phased-array approach, presented earlier in Eq. 6

$$y_m(t) = \sum_{n=1}^{N} w_{nm}s_n(t + \Delta_{nm}) \quad m = 1, 2 \cdots M$$

(6)

If the channel gains are all adjusted to be approximately equal, as we assumed for our derivation, then the weight coefficients $w_{n,m}$ reduce to $1/N$ for all channels, yielding Eq. 7 in our derivation

$$y_m(t) = \frac{1}{N} \sum_{n=1}^{N} s_n(t + \Delta_{nm}) \quad m = 1, 2 \cdots M$$

(7)

The resulting outputs $y_m(t)$ of the phased-array processing are also passed through a threshold detector, and those transients that surpass the threshold are identified as spikes from neuron $m$.

**Computational Implementation of the Two Methods.** A comparison of Eq. 19 and Eq. 7 shows how the phased-array approach enables a substantial computational advantage over...
the conventional filtering approach. For the phased-array method, temporal filters are not derived for the multichannel spike waveform templates, and no correlation computations are executed. Instead, a set of time-shifted delta functions are used to shift the different channels of data, to bring about the temporal alignment of specific units for subsequent averaging. One major consequence of the elimination of the filter convolution step in the phased-array approach is a substantial improvement in its computational speed over that of the filtering approach. The relative computational complexity of these two approaches can be assessed very easily. To do so, we consider only the aspects of the procedures that must be executed as part of the ongoing data analysis, and ignore the filter derivation step in the optimal filter approach because that part of the procedure can be done once at the initiation of the analysis (assuming subsequent stationarity). Assume that the optimal linear filters \( f_m(n) \) have been derived for template \( m \), for \( m = 1, 2, \cdots, M \), for all \( N \) channels from \( n = 1, 2, \cdots, N \). The calculation of the correlation between each filter and the data stream from all active channels is costly from a computational standpoint. Assuming the data sampling rate is \( F \), the sampling window size is \( W_t \), and \( K \) is the filter order, then the filtering calculation for discrimination of \( M \) templates from \( N \) channels of data, represented by Eq. 19, requires at the order of \( M \times N \times F \times W_t \times K \) multiplications and \( M \times N \times F \times W_t \times K \) summations.

The calculations required for the phased-array discrimination require many fewer operations. The procedure represented by Eq. 6 requires only at the order of \( M \times N \times F \times W_t \times K \) summation operations and \( M \times N \times F \times W_t \times K \) time-shifting operations. Further, the time-shifting operations can be achieved with much greater computational efficiency than multiplication operations, and can easily be implemented through hardware-based time-delay devices or through software time-delay operations. Thus the phased-array algorithm offers a huge computational advantage over linear filtering and is a good candidate for real-time implementation.

Robustness of the phased-array algorithm to reporting false-positives attributed to similar spike shapes

One further aspect of the phased-array algorithm that offers an advantage over the conventional filtering approach (and also over conventional template-matching approaches) is related to the manner in which the response from spurious transients is distributed within the analyzed output. As discussed in the text related to Fig. 2B, the phased-array processing of a section of multichannel data that contains an unmatched unit results in an extended “ripple” spread out around the time of occurrence of that background unit. This ripple is the typical, expected spurious transient from an incorrect unit: it is the result of summing \( N \) inappropriately shifted instances of the background unit, and dividing those nonoverlapping signals by \( N \). In other words, the signal energy resulting from this mismatch between the \( N \)-channel phased array and an incorrect unit is distributed along an extended period of the output signal, rather than being concentrated at a single time of occurrence of the background unit. In contrast, linear filtering and template-matching algorithms do, in fact, concentrate much more of the spurious signal energy resulting from convolution of filters (or templates) with mismatched spikes within a much shorter time span, thereby resulting in a transient with a higher peak than the corresponding ripple transient in the phased-array case. The reduced amplitude of spurious transients in the phased-array case offers a distinct advantage for threshold-based spike-discrimination operation.

Limitations of the phased-array technique

Despite the advantages of the phased-array algorithm discussed above, there are situations in which linear filtering and/or template-matching approaches will enable more robust spike discrimination. Specifically, consider a situation in which the spikes from 2 different neurons might have nearly identical propagation speeds, but have significantly different shapes. Because the phased-array approach is relatively insensitive to spike waveform, and depends primarily on differences in propagation speed, the spikes from these 2 cells would be indistinguishable from phased-array processing alone, and might certainly be discriminable through filtering or template matching.

However, we note that even in this case, implementation of phased-array analysis as a first processing stage might offer a substantial increase in computational efficiency. If the essential aspects of the 2 spike waveforms that allow their discrimination are maintained throughout the averaging process, a single-channel linear filtering (or template-matching) operation could be implemented on the output of the phased array matched to the propagation speed for those units. This would increase the necessary computation time by a single-channel correlation calculation, but would enable a reduction in the required computational time by a factor of \( 1/N \) compared to a conventional \( N \)-channel correlation.

Acknowledgments

We thank Drs. Charles M. Gray and Alex Dimitrov for helpful discussions.

Grants

This work was supported by National Institute of Mental Health Grant R01 MH-57179 and National Science Foundation Division of Experimental and Integrative Activities/Biological Information Technology and Systems Grant 0129895.

References


