Neuroscience 131 (2005) 1–11

QUANTITATIVE MEASURES OF CLUSTER QUALITY FOR USE IN EXTRACELLULAR RECORDINGS

N. SCHMITZER-TORBERT,[a] J. JACKSON,[a] D. HENZE,[b] K. HARRIS[c] AND A. D. REDISH[a]*

[a]*Department of Neuroscience, 6-145 Jackson Hall, 321 Church Street SE, University of Minnesota, Minneapolis MN 55455, USA*

[b]*Department of Neuroscience, Merck Research Laboratories, West Point, PA 19486-0004, USA*

[c]*Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, NJ 07102, USA*

Abstract—While the use of multi-channel electrodes (stereotrodes and tetrodes) has allowed for the simultaneous recording and identification of many neurons, quantitative measures of the quality of neurons in such recordings are lacking. In multi-channel recordings, each spike waveform is discriminated in a high-dimensional space, making traditional measures of unit quality inapplicable. We describe two measures of unit isolation quality, L_{ratio} and Isolation Distance, and evaluate their performance using simulations and tetrode recordings. Both measures quantified how well separated the spikes of one cluster (putative neuron) were from other spikes recorded simultaneously on the same multi-channel electrode. In simulations and tetrode recordings, both L_{ratio} and Isolation Distance discriminated well- and poorly-separated clusters. In data sets from the rodent hippocampus in which neurons were simultaneously recorded intracellularly and extracellularly, values of Isolation Distance and L_{ratio} were related to the correct identification of spikes.

Key words: extracellular recording, stereotrode, tetrode, multi-channel electrode, spike sorting, unit isolation.

A major question in neuroscience is how information is processed in single neurons and in larger ensembles of neurons. Extracellular recording techniques have allowed the direct examination of neural responsiveness by providing access to neural activity in intact, behaving animals. The extent to which the responses of a single neuron can be characterized depends on the ability to identify the action potentials originating from a single cell, and to discriminate these action potentials from other sources of electrical activity. In traditional microelectrode techniques, the signal-to-noise ratio (SNR) has served as a quantitative measure of unit quality: spikes with large SNR values are likely to represent spikes from a source located very near to the recording region of the electrode, and spikes with very large SNR are likely to come from a single cell (Lemon, 1984). In order to examine information processing in networks of neurons, it is often desirable to record multiple neurons simultaneously (Wilson and McNaughton, 1993; Redish et al., 2000; Harris, 2003; Rosenzweig et al., 2003; Brown et al., 2004; Bartho et al., 2004). However, when multiple units are detected on the same microelectrode, different cells are likely to be confused with one another.

This problem may be ameliorated by the use of multi-channel electrodes, such as stereotrodes (McNaughton et al., 1983), tetrodes (O'Keefe and Recce, 1993; Wilson and McNaughton, 1993), and silicon microelectrodes (Drake et al., 1988; Csicsvari et al., 2003). Depending on the physical relationship of neurons relative to the multi-channel electrode, the amplitude and extracellular waveform of a neuron on each channel will likely differ from that of a neuron in a different physical location (Lemon, 1984; Holt and Koch, 1999; Henze et al., 2000; Buzsáki, 2004). Spikes presumed to come from the same neuron will form clusters in a high dimensional feature space which can be separated from other clusters representing other simultaneously recorded cells and noise events. Tetrodes have been applied successfully to multiple brain structures, allowing for the simultaneous recording of large numbers of neurons in the rodent hippocampus (Wilson and McNaughton, 1993) and improving separation of single neurons in visual cortex when compared with single channel electrodes (Gray et al., 1995).

However, while the use of multi-channel electrodes has allowed for the simultaneous recording of large ensembles of neurons, there is no widely used quantitative measure of cluster quality comparable to SNR. Instead, when cluster quality is directly addressed, such measures are usually based on subjective estimates of how well segregated the spikes in a cluster are from other spikes recorded on the same electrode. These subjective estimates of cluster quality have serious drawbacks such as 1) they are highly dependent on the human observer, and 2) they are not likely to perform well at every point on the continuum of cluster quality from well-isolated clusters to poorly-isolated clusters. The development of quantitative cluster quality measures addresses these concerns by creating an objective metric for the evaluation of cluster quality. Reporting of such quantitative measures would allow for the better evaluation of experimental results and may improve the reproducibility of results across laboratories. There exists in the literature at least one quantitative method for evaluating cluster quality, proposed by Pouzat et al. (2002), which utilizes statistics of the noise distribution (i.e. non-spiking times) to evaluate unit quality. The two measures used in this paper have the advantage that

*Corresponding author. Tel: +1-612-626-3738; fax: +1-612-626-5009.
E-mail address: redish@ahc.umn.edu (A. D. Redish).
Abbreviations: SNR, signal-to-noise ratio.

0306-4522/05$30.00+0.00 © 2005 IBRO. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuroscience.2004.09.066

they are applicable even when no measurements of the noise distributions are available.

Two recently introduced measures, L_{ratio} (Schmitzer-Torbert and Redish, 2004) and Isolation Distance (Harris et al., 2001), quantify the quality of a cluster of extracellularly recorded spikes by calculating how well separated the spikes in the cluster are from other spikes recorded on the same multichannel electrode. Here, we use simulations and actual data to demonstrate the utility of these measures for identifying well-separated clusters. We apply these measures to data sets taken from the rodent hippocampus where hippocampal pyramidal cells were simultaneously recorded intracellularly and extracellularly with a tetrode (Henze et al., 2000), and to striatal (Schmitzer-Torbert and Redish, 2004) and hippocampal (J. Jackson and A. D. Redish, unpublished observations) recordings in which neurons were recorded extracellularly with tetrodes. Some of these results have been presented in abstract form (Jackson et al., 2003).

EXPERIMENTAL PROCEDURES

Cluster quality measures

In an extracellular recording, spike waveforms recorded on a single tetrode represent a mixture of spikes obtained from one or more cells and waveforms due to noise events, such as mechanical artifacts. Clustering is normally accomplished by calculating a set of features of each spike waveform, such as the amplitude on each channel of a tetrode. Spikes are then represented as points in a high-dimensional feature space and clusters are either identified by manual users using projections of this feature space or by automatic clustering methods. After clustering spikes into putative cells, it is important to ensure that spikes assigned to one cluster were well separated from other spikes recorded simultaneously. While cases in which clusters are quite well separated are easily identified by human observers (for instance, see Fig. 3 and 4), such approaches are highly dependent on the human observer, and are not likely to perform well at all points along a continuum of cluster quality. The primary aim of this paper is to evaluate two quantitative measures of cluster quality: L_{ratio} and Isolation Distance.

Each cluster divides the total data set into two mutually exclusive subsets: the set of *cluster spikes,* or spikes which are members of the cluster and thought to represent the activity of a single neuron, and the set of *noise spikes* which are spikes that are not members of the cluster and thought to represent other neurons and noise events. When the distribution of cluster spikes does not overlap with the distribution of noise spikes in a high-dimensional feature space, then it is likely that the division of the data into sets of cluster and noise spikes was appropriate (again, see Fig. 3 and 4). However, classification errors are expected in cases where the distributions have significant overlap (for instance, see Fig. 6). In this paper, we use the terms *well separated* to describe the case in which the distribution of cluster spikes has little or no overlap with the distribution of noise spikes and *poorly separated* to describe the case in which the distribution of cluster spikes has a large amount of overlap with the distribution of noise spikes.

For both L_{ratio} and Isolation Distance, calculations were performed in a high-dimensional feature space. Energy and the first principal component coefficient were calculated for each spike waveform recorded on each tetrode channel. Energy was defined as the square root of the sum of squares of each point in the waveform, divided by the number of samples in the waveform and was calculated as:

$$E_{i,j} = \frac{\|w_{i,j}\|}{n} = \frac{\sum_{k=1}^{n} w_{i,j,k}^2}{n} \quad (1)$$

where $E_{i,j}$ is the energy of the ith spike on the jth channel, $w_{i,j}$ is the extracellular waveform of spike i on channel j, n is the number of samples in the extracellular waveform $w_{i,j}$, and $w_{i,j,k}$ is sample k of the n sample extracellular waveform of spike i on channel j.

When calculating principal components, each waveform was normalized by its energy. This allows the principal components to be based on waveform shape rather than overall amplitude parameters. Principal components were determined separately for each tetrode recording.

The eight feature quantities (four tetrode channels×two features) defined each spike as a point in eight dimensional space. Both measures make use of a statistical quantity known as *Mahalanobis distance.* The Mahalanobis distance $D_{i,C}^2$ of spike i from the center of the cluster C is defined by the formula

$$D_{i,C}^2 = (x_i - \mu_C)^T \Sigma_C^{-1} (x_i - \mu_C) \quad (2)$$

where x_i is the feature vector for spike i, μ_C is the mean of the values of the spikes in cluster C, and Σ_C is the covariance matrix of the spikes in cluster C. The Mahalanobis distance allows for the measurement of the distance between points in a high-dimensional space where there exists correlation between dimensions. For instance, there is often a correlation between the peaks of spikes observed on each channel of the electrode. In such a case, the cluster takes on an elongated shape, reflecting the correlation between the features (see the clusters shown in Fig. 2–6 for examples of clusters with correlations in the energy observed on pairs of channels in a tetrode recording). Depending on where the noise spikes fall with respect to the long axis of the cluster in that two-dimensional space, the Euclidean distances from the center of the cluster to the noise spikes would not well represent how far noise spikes are from the cluster boundaries. The Mahalanobis distance has the effect of accounting for the correlations between these dimensions, and the distances calculated will reflect the location of spikes with respect to the center of the cluster, after rescaling the cluster spikes into a sphere. When all of the dimensions are uncorrelated, and the variances of each dimension are equal, the Mahalanobis distance is equivalent to the Euclidean distance.

L_{ratio}

L_{ratio} was originally described by Schmitzer-Torbert and Redish (2004) and applied to striatal data. If the distribution of cluster spikes is multivariate normal (i.e. Gaussian), then D^2 for cluster spikes will distribute as χ^2 with eight degrees of freedom (because the measure is performed in an eight dimensional feature space; D'Agostino and Stephens, 1986). The assumption of multivariate normality thus provides χ^2 with eight degrees of freedom as an expectation for the distribution of the D^2 values for cluster spikes.

A quantity L is calculated as:

$$L(C) = \sum_{i \notin C} 1 - CDF_{\chi^2_{df}}(D_{i,C}^2) \quad (3)$$

where $i \notin C$ is the set of spikes which are not members of the cluster and $CDF_{\chi^2_{df}}$ is the cumulative distribution function of the χ^2 distribution with df=8. Noise spikes which are close to the center of cluster C will contribute strongly to this sum, while noise spikes far from the center of cluster C will contribute little. A low value of L indicates that the cluster has a good "moat" and is well separated from other spikes recorded on the same tetrode. In contrast, a high value of L indicates that the cluster is not well separated, and is likely to both include spikes which are not part of the cluster and exclude spikes that are part of the cluster. The cluster quality measure, L_{ratio} was defined as L divided by the total number of spikes in the cluster.

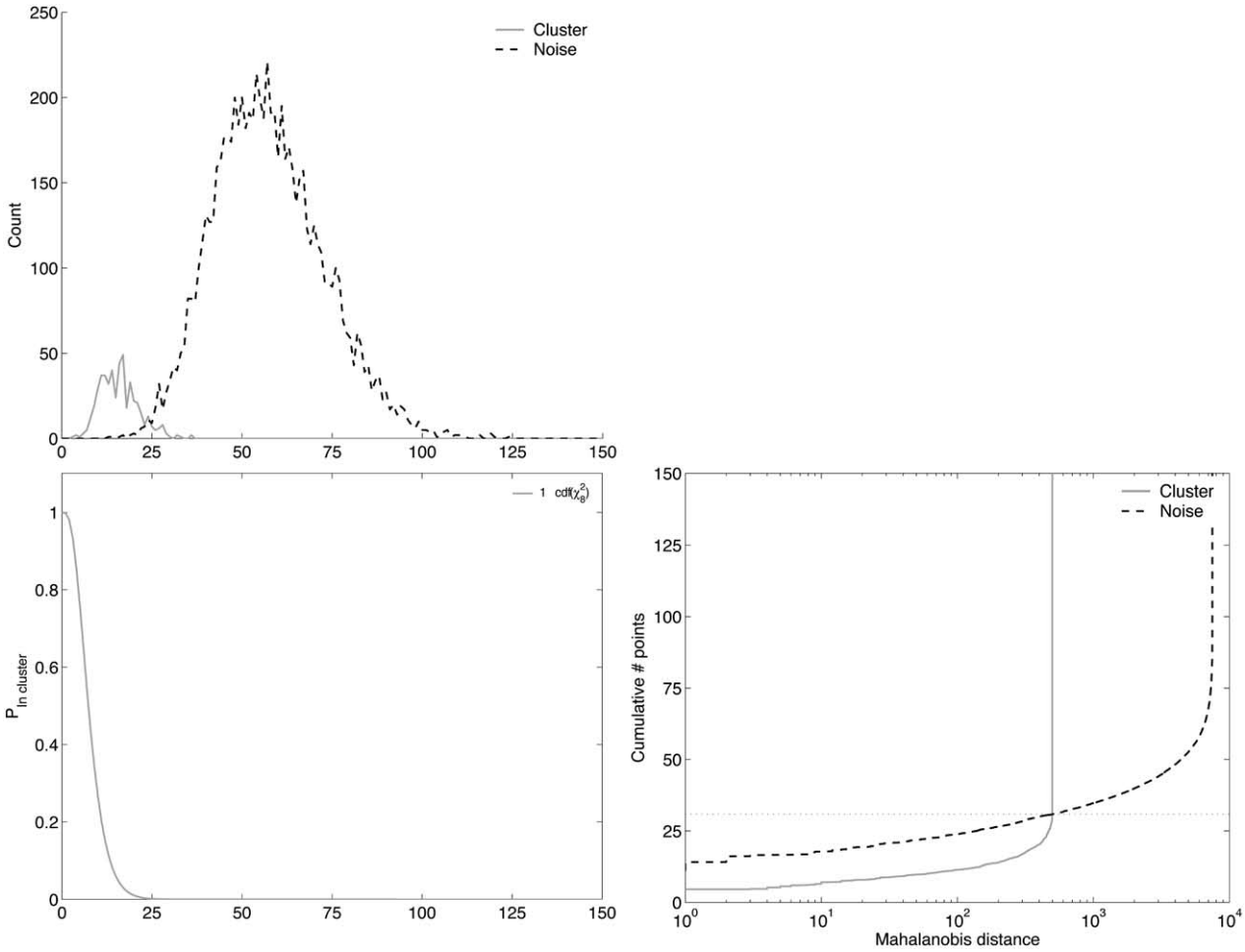

Fig. 1. Calculation of L_{ratio} (left) and Isolation Distance (right). For both plots, data points were based on a simulated noise and cluster distribution with unit variance. Both the cluster and noise were normally distributed in eight dimensions, and were separated by a distance of six standard deviations. Top: distributions of squared Mahalanobis distances (D^2) for the cluster and noise distributions. Solid lines indicate cluster points and dashed lines indicate noise points. Bottom left: L_{ratio}. Line indicates 1-cdf(χ_8^2). L_{ratio} is calculated by evaluating each of the noise points using the inverse of the χ^2 cdf, which indicates the probability that a noise point is near the center of the cluster. Bottom right: isolation distance. Cumulative count of points from a simulated cluster (solid line) and noise source (dashed line). This simulated cluster contains 500 points, and thus the isolation distance is the location of the 500th nearest noise point. When the cumulative count functions of the cluster spikes and noise spikes intersect exactly once, the value of Isolation Distance for the cluster is the intersection of the cumulative count functions.

$$L_{ratio}(C) = \frac{L(C)}{n_C} \tag{4}$$

where n_C is the number of spikes in C. Using a criterion based on L_{ratio} rather than L allows clusters with larger numbers of spikes to tolerate more contamination. L_{ratio} will be most suitable in cases where the responses of cells are being evaluated relative to some experimental parameter, as the effect of a given level of contamination on the observed tuning of the cell is likely to be proportional to the size of the cluster. However, in situations where even small amounts of contamination in large clusters are not acceptable, L may be a more appropriate quality measure than L_{ratio}.

Isolation Distance

Isolation Distance was first introduced by Harris et al. (2001) and applied to hippocampal data sets. If a cluster contains n_C cluster spikes, the Isolation Distance of the cluster is the D^2 value of the $n_C{}^{\text{th}}$ closest noise spike (see Fig. 1). Isolation Distance is therefore the radius of the smallest ellipsoid from the cluster center containing all of the cluster spikes and an equal number of noise spikes. As such, Isolation Distance estimates how distant the cluster spikes are from the other spikes recorded on the same electrode. Isolation Distance is not defined for cases in which the number of cluster spikes is greater than the number of noise spikes.

Simulations

To explore how Isolation Distance and L_{ratio} performed in a controlled context, each measure was tested using simulations in which a multivariate Gaussian distribution of spikes (the cluster) was separated by some distance from one or more sets of Gaussian-distributed noise spikes. The use of simulation data allowed us to examine the performance of each measure as a function of separation between a cluster and a noise source and the dimensionality of the feature space, as well as the performance in cases of multimodal noise. Simulations tested Gaussians with unit variance as well as non-Gaussian cluster distributions.

Neurophysiology

Data sets. To compare qualitative estimates of cluster quality to the performance of Isolation Distance and L_{ratio}, cluster examples from tetrode data collected from the rodent hippocampus (Henze et al., 2000) or dorsal striatum (Schmitzer-Torbert et al., 2002) were examined. For both hippocampal and striatal data sets, surgical and experimental protocols followed appropriate institutional guidelines for animal care. Experimental procedures were approved by the Institutional Animal Care and Use Committees of the appropriate institutions. These analyses were done on data collected for previous experiments; no additional animals were used.

Paired intracellular and extracellular data sets. Hippocampal data sets (Henze et al., 2000) were analyzed to determine the relationship between cluster quality and the correct identification of spikes from a cluster. Six paired recordings of one hippocampal pyramidal neuron recorded intracellularly and multiple neurons recorded extracellularly were used. The proportion of noise spikes incorrectly classified as cluster spikes and the proportion of cluster spikes incorrectly classified as noise spikes were examined for clusters created by manual users and an automatic clustering algorithm (for details, see Harris et al., 2000).

RESULTS

Simulations

In these simulations, both L_{ratio} and Isolation Distance discriminated between poorly-separated and well-separated clusters. When a cluster composed of 500 points was moved away from a noise distribution composed of 7500 points, both Isolation Distance and L_{ratio} differentiated poorly- and well-separated states. As the distance between the cluster and noise was increased, values of L_{ratio} and Isolation Distance improved.

As the dimensionality of the feature space was increased, values of Isolation Distance and L_{ratio} varied with the dimensionality. However, at each dimensionality, both measures differentiated poorly-separated clusters from well-separated clusters: all values of Isolation Distances from poorly-separated clusters were lower than those of well-separated clusters and all values of L_{ratio} from poorly-separated clusters were larger than those of well-separated clusters. These results indicate that both measures were effective in discriminating well- and poorly-separated clusters independent of the dimensionality of the feature space, but that any threshold used to define a well- or poorly-separated cluster will depend on the dimensionality of the feature space.

When the noise distribution was bimodal and a small noise mode was located near the center of the cluster, L_{ratio} outperformed Isolation Distance. Isolation Distance was not sensitive to the presence of the noise unless this small noise mode contained at least as many points as the cluster. L_{ratio} detected the presence of the noise points. These results would indicate that in cases where noise distributions are complex, L_{ratio} may provide a better estimate of the presence of local noise points.

Cellular spike clusters recorded on multi-channel electrodes are not always well described by Gaussian models due to spike-adaptation and other effects (Fee et al., 1996a,b; Lewicki, 1998; Harris et al., 2000; Shoham et al., 2003). Generally, clusters have longer tails than would be expected from a Gaussian distribution. Although this means that Gaussian models are not always well suited for automatic clustering, the question being addressed in this paper is the contamination of the cluster by noise (other neurons and non-neural noise). We tested multivariate *t*-distributions (Shoham et al., 2003) and mixtures of Gaussians (Lewicki, 1998) as these have been proposed as models of spike distributions. As expected, deviations from normality in these non-Gaussian distributions had no appreciable effect on the L_{ratio} or Isolation Distance measurements. This occurs because L_{ratio} and Isolation Distance measure the distribution of non-cluster points relative to the expected distribution of cluster points. Although a Gaussian distribution does not approximate multivariate *t*-distributions or mixtures-of-Gaussians distributions accurately enough to decide inclusion within a cluster, a Gaussian distribution does approximate them well enough to measure Mahalanobis distance to non-cluster points, and thus does approximate them well enough that L_{ratio} and Isolation Distance continue to provide robust measures of cluster quality.

We further applied these measures to simulations using actual hippocampal data sets. Fig. 2 shows an energy projection for data from a hippocampal tetrode recording. The spikes from an intracellularly identified hippocampal neuron are shown in black; all other points are shown in gray. As can be seen in the energy projections, this cluster was not well described by a Gaussian fit: the cluster had a bimodal distribution in this two-dimensional projection. Also shown in Fig. 2, L_{ratio} and Isolation Distance were calculated for the hippocampal neuron as the cluster was moved into and out of the noise distribution by multiplying the average waveform of the cluster by a scalar. In addition to L_{ratio} and Isolation Distance, the SNR was calculated for each tetrode channel for the cluster spikes (SNR defined as difference in the average energy of the cluster spikes from the average energy of the noise spikes, normalized by the standard deviation of the energy of the noise waveforms). As the cluster was moved out of the distribution of noise spikes, all measures improved. This demonstrates, using actual data, that L_{ratio} and Isolation Distance can successfully be used even with non-Gaussian clusters. SNR on each tetrode channel also improved as the cluster was moved out of the distribution of noise spikes, but did not provide an unambiguous measure of cluster quality.

Application to neural data

L_{ratio} and Isolation Distance agree with subjective estimates of cluster quality. Although L_{ratio} and Isolation Distance performed well on simulated data, it was also important to demonstrate that these measures discriminate clusters which are identified as poorly separated and well separated by human observers. Fig. 3–6 show four clusters taken from tetrode data collected in rat hippocampus or dorsal striatum. Fig. 3 shows data from a tetrode recording in the rodent hippocampus in which a well-separated cluster of spikes was observed. Fig. 4 shows a well-separated cluster from data recorded with a tetrode in

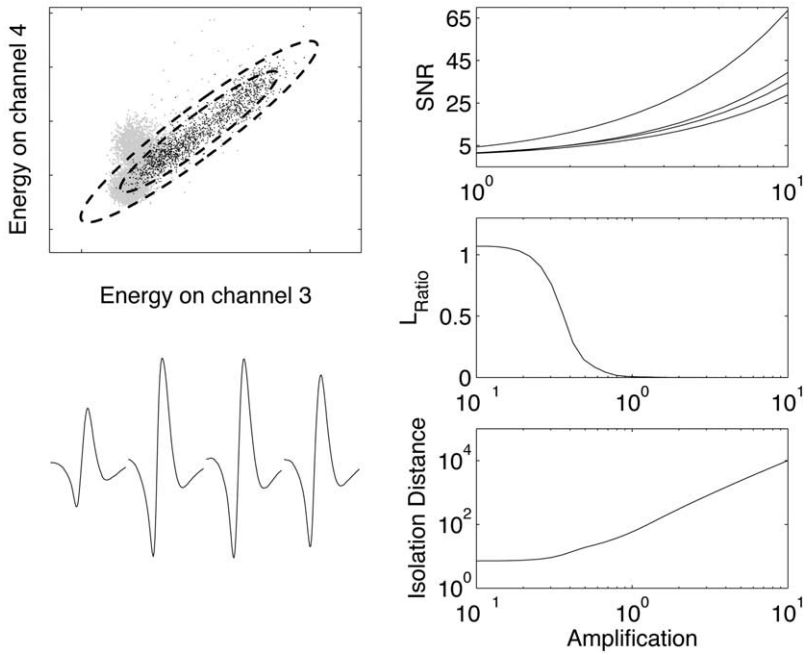

Fig. 2. Simulation with hippocampal data. Top left: actual data from a hippocampal cell. Spikes and noise events recorded on a single tetrode placed in the rodent hippocampus. Black points are spikes which were identified on the basis of intracellular recordings. Dashed lines indicate 2 and 3 standard deviation isocontours generated from a two-dimensional Gaussian fit of the cluster. Bottom left: average waveform on each tetrode channel of the hippocampal cell shown above. Right: changes in cluster quality as the center of the cluster is moved. The center of the cluster was shifted from what is shown in the top left panel by multiplying the average waveform of the cluster by a scalar and leaving the residuals intact. SNR for each of the four tetrode channels improved as the cluster was increasingly separated from the noise, but each channel gave a different SNR. Even using a highly bimodal cluster, L_{ratio} and Isolation Distance discriminated between situations in which the cluster was well and poorly separated.

the rodent striatum. The distributions of D^2 values for cluster spikes and noise spikes are also shown, as well as values of L_{ratio} and Isolation Distance for each cluster. D^2 was calculated in eight dimensions using the energy and the first principal component coefficient of the extracellular waveforms on each tetrode channel. For both clusters, there was very little overlap of the D^2 distributions of the cluster spikes and noise spikes, and values of L_{ratio} and Isolation Distance indicated that the clusters were well separated. Fig. 5 and 6 show two other clusters recorded

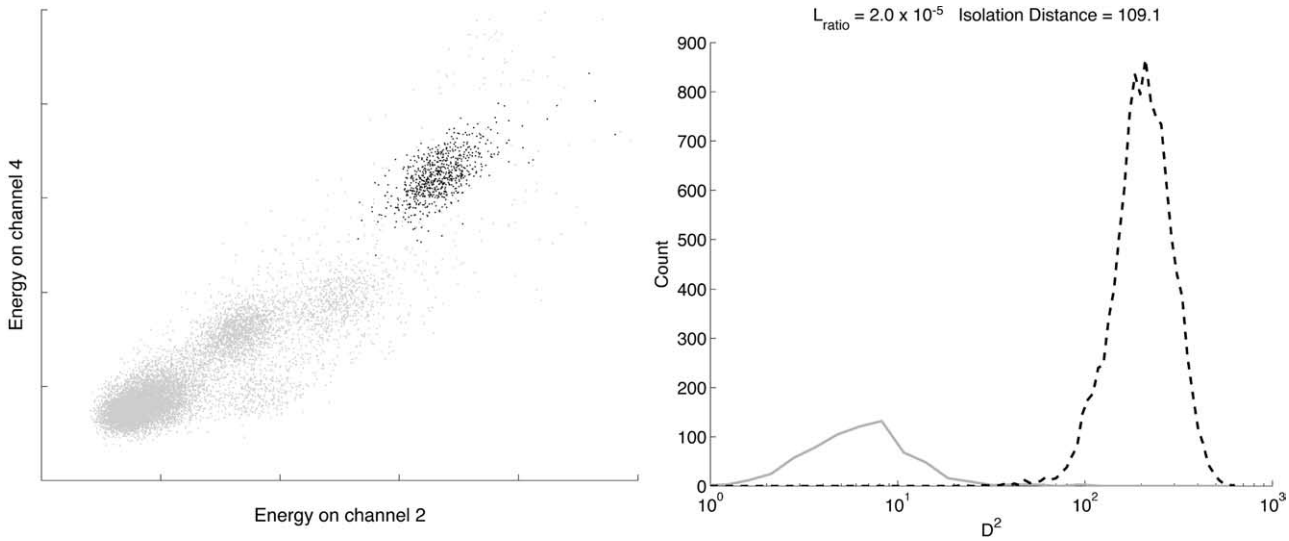

Fig. 3. Hippocampal cell with good separation. Left: spikes and noise events recorded on a single tetrode placed in the rodent hippocampus. Black points are spikes which were identified on the basis of intracellular recordings. Right: separation of the cluster from all other events recorded. L_{ratio} and Isolation Distance were calculated in eight dimensions. Solid line indicates the distribution of D^2 values for the cluster (black dots in left panel). Dotted line indicates the distribution of D^2 values for the non-cluster points (gray dots in left panel). (Data from Henze et al., 2000: 690 spikes in the cluster, 12,031 spikes recorded total).

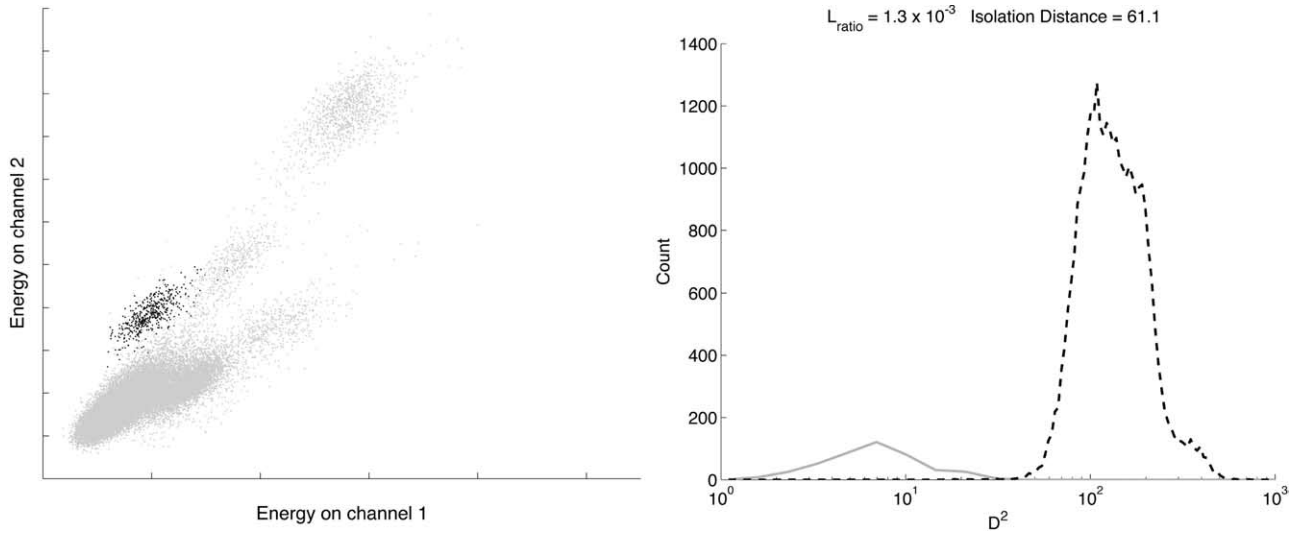

Fig. 4. Striatal cell with good separation. Left: spikes and noise events recorded on a single tetrode placed in the rodent striatum. Black points are spikes from a cluster defined using extracellular data. Right: separation of the cluster from all other events recorded. Format as per Fig. 3. (R023-2002-08-TT10-07: 436 spikes in the cluster, 31,394 spikes recorded total).

in rodent striatum. As can be seen in the energy projections and the D^2 distributions, the cluster spikes in Fig. 5 and 6 are not as well separated as the examples shown in Fig. 3 and 4. The worst separation is observed in Fig. 6, while Fig. 5 has an intermediate separation. Values of L_{ratio} and Isolation Distance agreed with this subjective categorization of cluster quality. The clusters shown in Fig. 3 and 4 had the best quantitative separation for both measures, Fig. 5 had an intermediate value, and Fig. 6 had the worst values of both L_{ratio} and Isolation Distance. These examples support the use of these cluster quality measures as an objective method for evaluating cluster quality.

Comparison of cluster quality from separate tetrodes. A highly desirable feature of cluster quality measurements is that they are comparable across separate recording sessions. As cluster quality values of L_{ratio} and Isolation Distance depend on the dimensionality of the feature space (described above), comparisons across sessions must be made between feature spaces of equal dimensionality. As different waveform features (e.g. energy, peak amplitude, principal component coefficients) can be correlated with one another (e.g. the peak and energy of a waveform are not independent), these correlations between features reduce the effective dimensionality of a given feature space. We stress that direct comparisons of cluster qualities calculated using different feature spaces are not justified. For example, cluster qualities should not be compared directly between clusters evaluated with four-dimensional versus

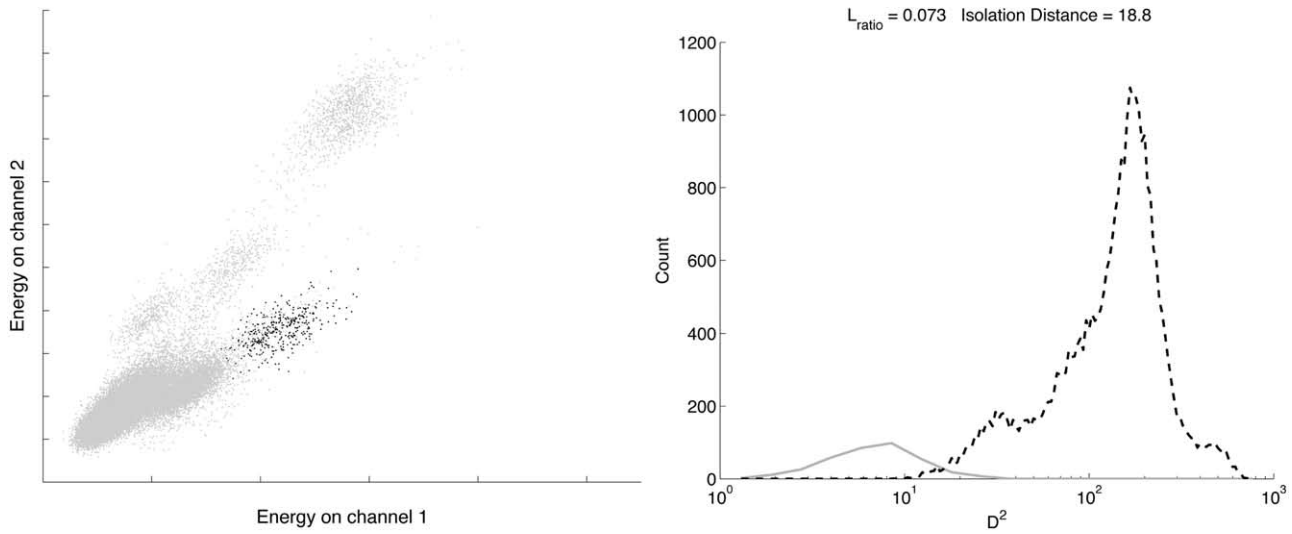

Fig. 5. Striatal cell with intermediate separation. Left: spikes and noise events recorded on a single tetrode placed in the rodent striatum. Black points are spikes from a cluster defined using extracellular data. Right: separation of the cluster from all other events recorded. Format as per Fig. 3. (R023-2002-08-TT10-05: 360 spikes in the cluster, 31,394 spikes recorded total).

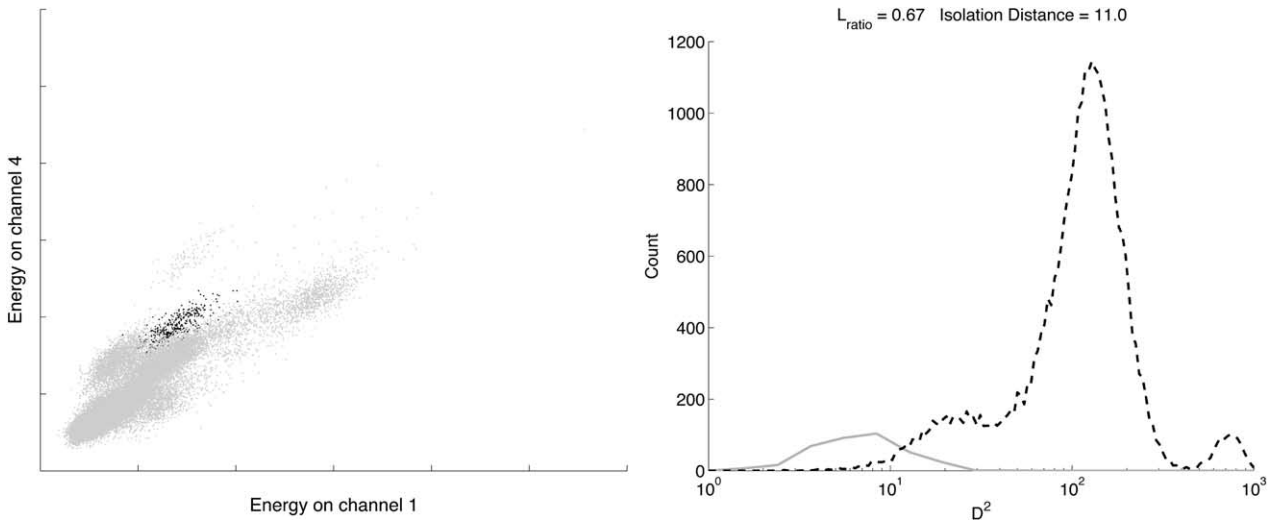

Fig. 6. Striatal cell with poor separation. Left: spikes and noise events recorded on a single tetrode placed in the rodent striatum. Black points are spikes from a cluster defined using extracellular data. Right: separation of the cluster from all other events recorded. Format as per Fig. 3. (R023-2002-08-TT10-08: 367 spikes in the cluster, 31,394 spikes recorded total).

eight-dimensional feature spaces, or between clusters evaluated with two eight-dimensional feature spaces created using different waveform features (energy/first principal component coefficients versus peak amplitude/second principal component coefficients). Thus, cluster quality comparisons made between clusters recorded in different recording sessions should be performed using feature spaces defined by the same set of extracellular waveform features, and the same number of electrode channels. (This does not mean that clusters must be created using the same feature space which is used to calculate cluster quality. Clusters can be created by any method deemed appropriate (manual cluster cutting, automatic spike sorting, etc.). Then, after clusters have been created from a data set, cluster qualities must be calculated on the basis of a fixed, or standardized, feature space in order that cluster qualities can be compared across separate recordings.)

A secondary concern relates to the use of feature spaces which include principal component coefficients of the extracellular waveforms. The cannonical vectors used to compute the principal component coefficients are dependent on the distribution of waveforms obtained on the tetrode and thus can change between recording sessions. These changes in the cannonical vectors might have an effect on the values of both L_{ratio} and Isolation Distance, which could limit our ability to make comparisons between the quality of clusters from separate recording sessions. We estimated the variability in values of L_{ratio} and Isolation Distance by recomputing cluster quality values using principal component vectors taken from separate tetrodes. Across a set of 55 tetrode recordings from the striatum (27 recordings, 117 spike trains) and hippocampus (28 recordings, 145 spike trains), the median absolute error in L_{ratio} and Isolation Distance on the log scale was 0.071 and 0.023 respectively. In these samples, the average L_{ratio} and Isolation Distance values on the log scale were -1.52 ± 1.48 and 1.44 ± 0.37 respectively (mean $\pm$ standard deviation). As the error rates for each measure were much smaller than the width of the distribution of cluster quality values, we can conclude that the amount of error introduced by using the within-tetrode principal components is negligible.

Paired intracellular and extracellular data

Type I and type II errors. Although values of L_{ratio} and Isolation Distance agreed with qualitative evaluations of well-separated and poorly separated clusters for the tetrode data shown above, a better test of the utility of each measure is the relationship between values of L_{ratio} and Isolation Distance to the correct identification of the spikes actually originating from one neuron. Because such information is not available in an extracellular recording, we analyzed hippocampal data sets in which a hippocampal neuron was simultaneously recorded both intracellularly and extracellularly (Henze et al., 2000). In these data, the identity of each spike from the neuron in the extracellular data can be identified by the intracellular recording. Classification errors for clusters created using the extracellular data were compared with values of L_{ratio} and Isolation Distance for six hippocampal neurons. For each cluster, the proportion of noise spikes which were incorrectly included in the cluster (type I error) and the proportion of cluster spikes which were incorrectly excluded from the cluster (type II error) were examined separately.

As shown in Fig. 7, both L_{ratio} and Isolation Distance were related linearly on the log scale to type I and type II error rates. While both measures were significantly correlated to both error types, the strongest relationships were between L_{ratio} and type II (false negative) error rates, and between Isolation Distance and type I (false positive) error rates. This suggests that these measures may be used independently as estimates of the two error types.

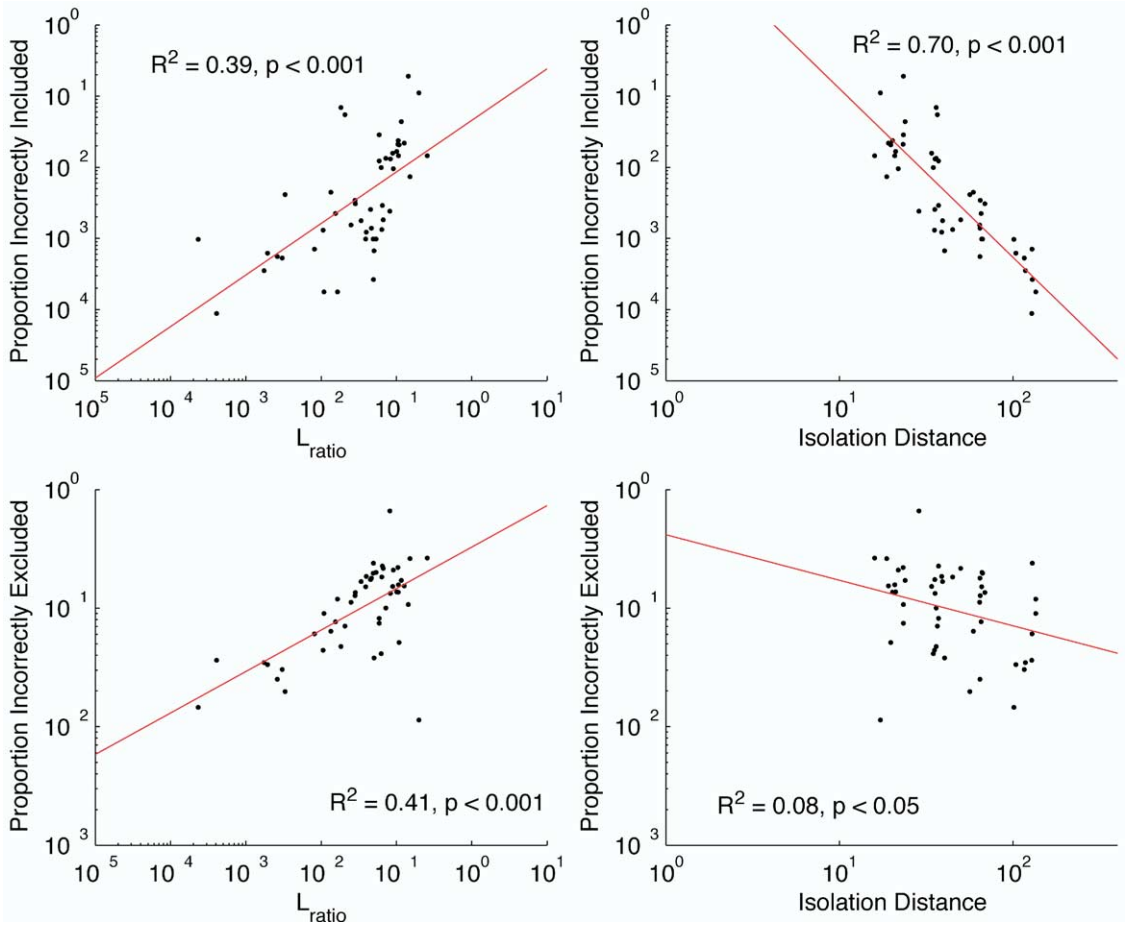

Fig. 7. Relationship between cluster quality and the correct identification of cell spikes. Cluster quality values for six cells identified by paired intra- and extracellular recordings. L_{ratio} (left) and Isolation Distance (right) values were calculated using clusters created by manual users and one automatic clustering technique (for details, see Harris et al., 2000). For each cluster, the proportion of spikes incorrectly included is shown at top, and the proportion of spikes incorrectly excluded by the same users is shown at bottom. For both L_{ratio} and Isolation Distance, decreases in cluster quality were associated with increases in the proportion of spikes incorrectly classified. For this data set, L_{ratio} was well related to the proportion of cluster spikes incorrectly excluded from the cluster (type II error), while the Isolation Distance was well related to the proportion of noise spikes incorrectly included in the cluster (type I error). All four correlations were significant.

Choice of features. The relationships between cluster quality measures and error rates shown in Fig. 7 held true over a wide range of feature spaces. To measure how the relationship between cluster quality values and error rates depended on the choice of waveform features, we repeated the analysis of the paired intracellular/extracellular tetrode data sets with nine commonly used waveform features (energy, peak amplitude, area [sum of the absolute value of the waveform samples], first, second and third principal component coefficients of the raw extracellular waveform, and first, second and third principal component coefficients of the energy normalized waveform). There was a significant improvement in the correlations of cluster quality values with error rates ($F_{(1,178)}=9.68$, $P=0.002$) when using feature spaces defined by pairs of features (two features measured on each tetrode channel: eight-dimensional feature spaces) relative to feature spaces defined by a single feature (one feature measured on each tetrode channel: four-dimensional feature spaces). While even larger feature spaces (using three or more features) might yield further improvements in the correlations between cluster quality and error rates, further analyses were limited to pairs of features in order to minimize the complexity of application of the cluster quality measures. The results for pairs of features (eight-dimensional feature spaces) held true for even higher dimensional feature spaces defined by three features (12-dimensional feature spaces) and are presumed to hold for higher dimensional spaces as well.

L_{ratio} and Isolation Distance were calculated using feature spaces defined by pairs of two waveform features calculated for each tetrode channel (a total of 36 eight-dimensional feature spaces, including the feature space used in Fig. 7). For each eight-dimensional feature space, the correlations of each cluster quality measure with type I and type II error rates were calculated on the log scale. As shown in Fig. 8, across the set of feature pairs, L_{ratio} and Isolation Distance were well correlated with type I and type II errors. In a two factor ANOVA (Measure×Error Type), there was a significant Measure×Error Type interaction

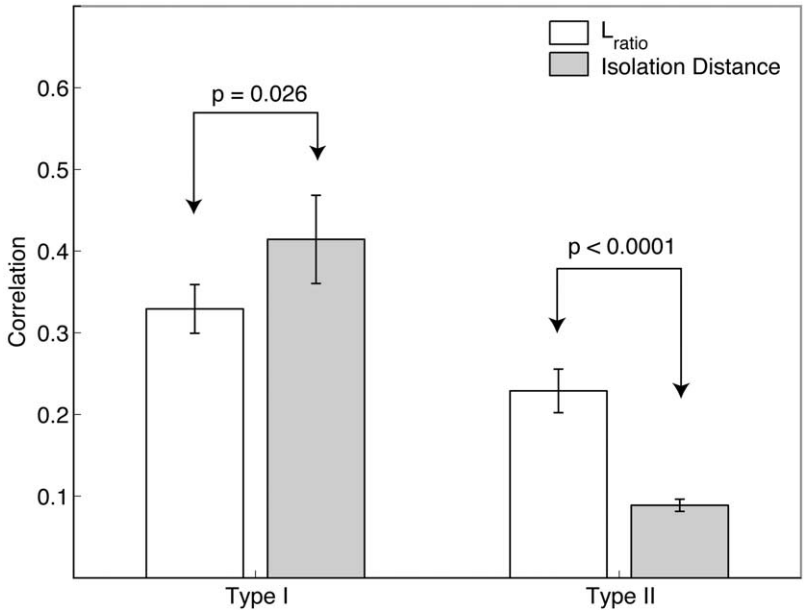

Fig. 8. Over a wide range of feature choices, L_{ratio} and Isolation Distance are well related to the number of type I and type II errors. For the paired hippocampal intracellular/extracellular data shown in Fig. 7, L_{ratio} and Isolation Distance were calculated using pairs of waveform features, and correlations with cluster quality values with the type I and type II error rates were calculated. Cluster quality values were calculated using feature spaces defined by all possible pairs of nine waveform features (energy, peak amplitude, area [sum of the absolute value of the waveform], First, second and third principal component coefficients of the raw extracellular waveform, and first, second and third principal component coefficients of the energy normalized waveform), for a total of 36 feature spaces. Bars represent the mean correlation of each cluster quality measure with type I and type II errors over the 36 feature spaces. Error bars indicate 95% confidence intervals. As with cluster quality values calculated using energy and the first principal component coefficient of the energy normalized waveform (Fig. 7), type I and type II error rates were correlated with both cluster quality measures over a wide range of feature spaces. On average, Isolation Distance was more highly correlated with type I error rates than was L_{ratio}, while L_{ratio} was more highly correlated with type II error rates than was Isolation Distance. Both measures were better correlated with type I than type II error rates.

($F_{(1,140)}$=45.7, P<0.0001). Post hoc tests (Tukey-Kramer) revealed that Isolation Distance was more highly correlated with type I errors than was L_{ratio}, while L_{ratio} was more highly correlated with type II errors than was Isolation Distance. There was also a main effect of Error Type ($F_{(1,140)}$=163.7, P<0.0001), indicating that both L_{ratio} and Isolation Distance had higher correlations with type I errors than type II errors. There was no main effect of Measure ($F_{(1,140)}$=2.7, P=0.102).

The correlation between error rates and cluster quality measures held over many feature spaces, indicating that a large set of waveform features is adequate for assessing cluster quality. However, the use of different feature spaces for reporting cluster quality values in published data would limit comparisons of cluster quality between publications. Therefore, a common set of features should be adopted from the set of acceptable features and used in calculating cluster quality values for published reports. Across the 36 feature spaces in which the paired intracellular/extracellular tetrode data sets were examined, the strongest correlations between cluster quality and error rates were obtained using the energy and the first principal component coefficient of the energy normalized waveform. As these features are also commonly used measures of extracellular waveforms in processing multi-electrode data, we recommend using a feature space defined by energy and the first principal component coefficient of the energy normalized waveform for reporting published cluster quality values.

DISCUSSION

While multichannel electrodes have allowed for better isolation of cells on the basis of extracellularly recorded action potentials, there exists a need for quantitative assessment of cluster quality. Two such quantitative measures, L_{ratio} and Isolation Distance evaluate the separation of a cluster from other spikes recorded on the same electrode. L_{ratio} is a measure of the amount of noise observed in the vicinity of the cluster, and Isolation Distance is a measure of how distant a cluster is from the noise distribution. On the basis of simulation, subjective evaluation and the correct identification of intracellularly recorded spikes, L_{ratio} and Isolation Distance performed well at quantifying cluster quality. Using data sets in which hippocampal pyramidal cells were simultaneously recorded intracellularly and extracellularly, we have shown that these cluster quality measures have a relationship with the percentage of spikes correctly identified in the extracellular recordings.

In simulations, L_{ratio} and Isolation Distance differentiated well-separated clusters from poorly-separated clusters. Also, while the values of Isolation Distance and L_{ratio} varied with the dimensionality of the feature space, at each dimensionality well-separated and poorly separated clusters were differentiated by both measures. These simula-

tions do not define what threshold should be used to define a well-separated cluster, but do indicate that these measures will provide quantitative measures of cluster quality.

Based on the correct identification of intracellularly recorded spikes, both Isolation Distance and L_{ratio} were related to the quality of extracellular spike classification. In these data sets, L_{ratio} was better related to the proportion of intracellular spikes which were "missed" while Isolation Distance was better related to the proportion of noise spikes which were incorrectly classified as part of the cluster. These data demonstrate the usefulness of both measures for evaluating the quality of clusters created solely based on extracellular data.

The strong correlations of Isolation Distance and L_{ratio} with type I and type II error rates, respectively, may arise from the specifics of what they measure. Isolation distance estimates the distance from the cluster to the nearest surrounding clusters; clusters that are close to their neighbors are more likely to contain contaminating spikes. L_{ratio} estimates the number of non-cluster spikes that lie in the "moat" immediately outside the cluster boundary; clusters with many nearby non-cluster spikes are more likely to have missed spikes which were actually generated from the cell in question.

Suggested application to neural data

The cluster quality measures L_{ratio} and Isolation Distance are not a replacement for the data processing used to identify cells in extracellular recordings. Rather, these cluster quality measures will assist neurophysiology research by providing objective criteria for the inclusion of units in further data analyses.

The suggested method for applying L_{ratio} and Isolation Distance is as follows: first, the extracellularly recorded action potentials should be sorted into putative clusters. Second, for each spike, some set of feature parameters must be calculated. The cluster quality measures are applicable to any such set of features, including peak amplitude, and peak-to-valley measurements, but for consistency between published reports, we suggest that published cluster quality values use the energy and first principal component coefficient of the energy normalized waveform. Third, once features have been calculated, the cluster quality of each cell can be calculated.

For the purposes of using these cluster quality values in published data, there are at least two ways to incorporate these measures into analyses: 1) the use of a threshold to define the minimum value of each cluster quality measure in order for a cell to be considered in further analyses, or 2) the description of the distribution of cluster quality values obtained. The use of a minimum threshold for cluster quality is required in cases where the information encoded by single cells is critical, for instance in cases where the tuning of individual neurons is examined relative to sensory or behavioral parameters.

The choice of threshold may itself depend on the scientific question being asked. In general, a good strategy is to compute the quantity of interest for all cells, and plot the dependency of this quantity against isolation quality. If dependence on cluster quality is seen, poorly isolated neurons should be excluded from further analysis. The appropriate threshold should be taken as the value above which no further dependence on cluster quality is observed (Harris, 2003; on-line supplementary material).

In other cases, it may be acceptable to include every cell in an analysis, for instance in cases of the reconstruction of behavioral or stimulus parameters from neural data (Georgopoulos et al., 1983; Jensen and Lisman, 2000; Johnson et al., in press; Salinas and Abbott, 1994; Wilson and McNaughton, 1993; Zhang et al., 1998). In this case, we recommend that descriptive statistics should be reported of the distribution of cluster qualities (i.e. mean and standard deviation), to enable objective evaluation of extracellular recording. In any report including these cluster quality measures, it is important to clearly describe how features were calculated, in order that results from different labs can more directly be compared. We would further suggest that the features used for published cluster quality values be the energy and first principal component coefficient of the energy normalized waveform.

Conclusions

Multi-channel recordings have allowed for the better isolation of extracellularly recorded cells with lower signal-to-noise ratios than single-electrode techniques. While the signal-to-noise ratio has served as a measure of unit quality in single-channel recordings, few quantitative methods have been described for multi-channel data. Isolation Distance and L_{ratio} offer a significant advance to multi-channel recordings by providing a quantitative method for evaluating cluster quality. Wider use of quantitative measures of cluster quality would likely improve the reproducibility of results across laboratories, and reporting of such quantitative measures would allow for the better evaluation of experimental results.

These two measures provide a general, quantitative method with which to gauge the contamination of a cluster in a high-dimensional feature space. As such, they are not limited to tetrode recordings. Separating cells from noise is required with any multi-channel electrode, including stereotrodes (McNaughton et al., 1983), tetrodes (O'Keefe and Recce, 1993; Wilson and McNaughton, 1993), and silicon microelectrodes (Drake et al., 1988; Csicsvari et al., 2003; Buzsáki, 2004; Bartho et al., 2004). Because signals occur on multiple channels simultaneously, standard measures of unit quality (such as SNR; Lemon, 1984) are inappropriate. L_{ratio} and Isolation Distance provide measures of cluster quality and as such, can provide measures of cell isolation quality for any multi-channel electrode, or single-channel electrode for which multiple features are considered (i.e. energy, principal component coefficients, etc.).

While Isolation Distance and L_{ratio} are quantitative measures of cluster quality, the choice of acceptable values of each measure will still depend on the experimental question being addressed. However, by reporting the minimum acceptable values of Isolation Distance and L_{ratio} obtained for a data set, other researchers would have

some idea about how much contamination was present. As such, reporting of Isolation Distance and L_{ratio} values would be a great improvement over having little or no information about the quality of the cells described.

Acknowledgments—We would like to thank G. Buzsaki for generously supplying the simultaneous intracellular/extracellular hippocampal data. We would also like to thank Matthew Wiener for helpful comments on an earlier draft of this manuscript. This work was partially supported by NSF-IGERT 9870633 (N.C.S.-T., J.C.J.), by an NSF fellowship to N.C.S.-T. by NIH MH68029 (A.D.R., J.C.J.) and by NIH MH73245 (K.D.H.).

REFERENCES

Bartho P, Hirase H, Monconduit L, Zugaro M, Harris KD, Buzsaki G (2004) Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. J. Neurophysiol 92:600–608.

Brown EN, Kass RE, Mitra PP (2004) Multiple neural spike train data analysis: state-of-the-art and future challenges. Nat Neurosci 7:456–461.

Buzsáki G (2004) Large-scale recording of neuronal ensembles. Nat Neurosci 7:446–451.

Csicsvari J, Henze DA, Jamieson B, Harris KD, Sirota A, Bartho P, Wise KD, Buzsaki G (2003) Massively parallel recording of unit and local field potentials with silicon-based electrodes. J Neurophysiol 90:1314–1323.

D'Agostino RB, Stephens MA, editors (1986) Goodness-of-Fit Techniques. New York: Marcel Dekker.

Drake KL, Wise KD, Farraye J, Anderson DJ, BeMent SL (1988) Performance of planar multisite microprobes recording extracellular single-unit intracortical activity. IEEE Transact Biomed Engineer 35:719–732.

Fee MS, Mitra PP, Kleinfeld D (1996a) Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. J Neurosci Methods 69:175–188.

Fee MS, Mitra PP, Kleinfeld D (1996b) Variability of extracellular spike waveforms of cortical neurons. J Neurophysiol 76:3823–3831.

Georgopoulos AP, Caminiti R, Kalaska JF, Massey JT (1983) Spatial coding of movement: a hypothesis concerning the coding of movement direction by motor cortical populations. Exp Brain Res Suppl: 327–336.

Gray CM, Maldonado PE, Wilson M, McNaughton BL (1995) Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings on cat striate cortex. J Neurosci Methods 63:43–54.

Harris KD (2003) Organization of cell assemblies in the hippocampus. Nature 424:552–556.

Harris KD, Henze DA, Csicsivari J, Hirase H, Buzsaki G (2000) Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. J Neurophysiol 84: 401–414.

Harris KD, Hirase H, Leinekugel X, Henze DA, and Buzsáki G (2001) Temporal interaction between single spikes and complex spike bursts in hippocampal pyramidal cells. Neuron 32:141–149.

Henze DA, Borhegyi Z, Csicsvari J, Mamiya A, Harris KD, Buzsaki G (2000) Intracellular features predicted by extracellular recordings in the hippocampus *in vivo*. J Neurophysiol 84:390–400.

Holt GR, Koch C (1999) Electrical interactions via the extracellular potential near cell bodies. J Comput Neurosci 6:169–184.

Jackson J, Schmitzer-Torbert NC, Harris KD, Redish AD (2003) Quantitative assessment of extracellular multichannel recording quality using measures of cluster separation. Soc Neurosci Abstr 518.18.

Jensen O, Lisman JE (2000) Position reconstruction from an ensemble of hippocampal place cells: contribution of theta phase encoding. J Neurophysiol 83:2602–2609.

Johnson A, Seeland KD, Redish AD. Reconstruction of the postsubiculum head direction signal from neural ensembles. Hippocampus, in press.

Lemon R (1984) Methods for neuronal recording in conscious animals. IBRO handbook series: methods in the Neurosciences. New York: Wiley.

Lewicki MS (1998) A review of methods for spike sorting: the detection and classification of neural action potentials. Network: Comput Neural Syst 9:R53–R78.

McNaughton BL, O'Keefe J, Barnes CA (1983) The stereotrode: a new technique for simultaneous isolation of several single units in the central nervous system from multiple unit records. J Neurosci Methods 8:391–397.

O'Keefe J, Recce M (1993) Phase relationship between hippocampal place units and the EEG theta rhythm. Hippocampus 3:317–330.

Pouzat C, Mazor O, Laurent G (2002) Using noise signature to optimize spike-sorting and to assess neuronal classification quality. J Neurosci Methods 122:43–57.

Redish AD, Rosenzweig ES, Bohanick JD, McNaughton BL, Barnes CA (2000) Dynamics of hippocampal ensemble realignment: time vs. space. J Neurosci 20:9289–9309.

Rosenzweig ES, Redish AD, McNaughton BL, Barnes CA (2003) Hippocampal map realignment and spatial learning. Nat Neurosci 6:609–615.

Salinas E, Abbott L (1994) Vector reconstruction from firing rates. J Comput Neurosci 1:89–107.

Schmitzer-Torbert NC, Redish AD (2004) Neuronal activity in the rodent dorsal striatum in sequential navigation: separation of spatial and reward responses on the multiple-T task. J Neurophysiol 91:2259–2272.

Schmitzer-Torbert NC, Jackson JC, Redish AD (2002) Behavioral correlates of neuronal activity in the rodent dorsal striatum: the multiple-T task. Soc Neurosci Abstr 676.4.

Shoham S, Fellows MR, Normann RA (2003) Robust, automatic spike sorting using mixtures of multivariate t-distributions. J Neurosci Methods 127:111–122.

Wilson MA, McNaughton BL (1993) Dynamics of the hippocampal ensemble code for space. Science 261:1055–1058.

Zhang K, Ginzburg I, McNaughton BL, Sejnowski TJ (1998) Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. J Neurophysiol 79:1017–1044.

(Accepted 1 September 2004)