Wavelet-based processing of neuronal spike trains prior to discriminant analysis

Mark Laubach*

John B. Pierce Laboratory and Department of Neurobiology, Yale University, 290 Congress Ave, New Haven, CT 06519, USA

Abstract

Investigations of neural coding in many brain systems have focused on the role of spike rate and timing as two means of encoding information within a spike train. Recently, statistical pattern recognition methods, such as linear discriminant analysis (LDA), have emerged as a standard approach for examining neural codes. These methods work well when data sets are over-determined (i.e., there are more observations than predictor variables). But this is not always the case in many experimental data sets. One way to reduce the number of predictor variables is to preprocess data prior to classification. Here, a wavelet-based method is described for preprocessing spike trains. The method is based on the discriminant pursuit (DP) algorithm of Buckheit and Donoho [Proc. SPIE 2569 (1995) 540–51]. DP extracts a reduced set of features that are well localized in the time and frequency domains and that can be subsequently analyzed with statistical classifiers. DP is illustrated using neuronal spike trains recorded in the motor cortex of an awake, behaving rat [Laubach et al. Nature 405 (2000) 567–71]. In addition, simulated spike trains that differed only in the timing of spikes are used to show that DP outperforms another method for preprocessing spike trains, principal component analysis (PCA) [Richmond and Optican J. Neurophysiol. 57 (1987) 147–61].
© 2003 Elsevier B.V. All rights reserved.

Keywords: Spike train; Neural coding; Discriminant analysis; Pattern recognition; Feature extraction; Preprocessing; Dimension reduction; Wavelets

1. Introduction

A major unresolved issue in systems neuroscience is the extent to which neurons make use of variations in spike rate and timing to transmit information (Engel et al., 1992; Shadlen and Newsome, 1995; Softky, 1995; Theunissen and Miller, 1995). A variety of data analysis paradigms have been developed to examine this issue, including multivariate statistical methods such as discriminant analysis (Miller et al., 1991; Gochin et al., 1994; Schoenbaum and Eichenbaum, 1995; Deadwyler et al., 1996; Nicolelis et al., 1997b, 1998; Ghazanfar et al., 2000; Laubach et al., 2000; Furukawa and Middlebrooks, 2002). Methods for discriminant analysis generate statistical models that predict whether a given class of signal occurred on a single trial. The results of a given discriminant analysis can easily be converted to the terms and metrics of information theory (i.e., bits of information). For spike train data, methods for discriminant analysis perform well when relatively large bin sizes

(>100 ms) are used as inputs for a given statistical classifier; however, the use of smaller bin sizes (<10 ms) tends to result in degraded performance (Nicolelis et al., 1999). This is likely due to a well-known aspect of algorithms such as linear discriminant analysis (LDA; Fisher, 1936) and other methods based on artificial neural networks (ANNs; see Ripley, 1996; Hastie et al., 2001 for review): the data set must be over-determined, i.e., there needs to be many more observations than predictor variables. The development of newer methods for pattern recognition, such as support vector machines (SVMs) (Vapnik, 2000; see Hastie et al., 2001 for review), may help overcome this issue. Nevertheless, to resolve whether a given set of neurons make use of variations in spike rate or timing, it is necessary to examine the neuronal response over a set of small bins. Reducing the dimensionality of spike train data prior to analysis with statistical classifiers is thus a major goal for current research.

A standard multivariate approach to dimension reduction is to use principal component analysis (PCA). PCA reduces a large number of original variables into a smaller number of "components" that represent most of the variance in the original data. Richmond and Optican (1987) developed an

---

* Tel.: +1-203-562-9901; fax: +1-203-624-4950.
E-mail address: mark.laubach@yale.edu (M. Laubach).

application of PCA for spike train data in the 1980s. The method revealed that visual cortical neurons make use of more than just firing rate to encode information about a collection of stimuli. In Richmond and Optican's method, each principal component is formed by the weighted linear sum of the firing rate of the neuron over the collection of bins that form the peri-event histogram. Coefficients for each component are determined for each bin in the peri-event histogram and each principal component is derived in an orthogonal manner (i.e., each accounts for a unique portion of variance). By taking the dot product of the coefficients and the observed spike train, one can compute an overall score for the component on a single-trial basis. The collection of scores for all trials in a given experimental data set can then be analyzed with statistical classifiers (Heller et al., 1995) or information theory (Richmond and Optican, 1987).

PCA is known to have limited utility when the signals of interest are sparsely distributed, e.g., the difference between firing patterns is based on one or a small number of spikes occurring within a narrow time window. This issue is due to an emphasis by PCA on global features in signals and images (Penev and Atick, 1996) that compromises the resolution of localized features. Alternatives to PCA have been developed since the early 1990s, e.g., independent component analysis (ICA) (see Hyvarinen et al., 2001 for review) and multi-scale spectral methods such as wavelets (see Wickerhauser, 1994 for review). These methods are not compromised by the sparseness of data and, in the case of wavelets, can account for structure in both the time and frequency domains.

Saito and Coifman (1994) described a method called local discriminant bases (LDB) that can be used to discriminate between multiple classes of signals by using expansions of the data in the wavelet (or local trigonometric) domain. Each signal in a given training data set is decomposed into a wavelet packet table, which is a way of describing the complete, orthonormal set of basis functions for a given wavelet function. Features derived with LDB are mutually perpendicular, i.e., their dot product is thus equal to zero, and are well localized in time and frequency. A criterion function that can be used for classification problems is, for example, the relative entropy (i.e., the Kullback–Leibler divergence). A subset of components with the maximum relative entropy are chosen as a set of features and used as input for statistical classifiers.

Buckheit and Donoho (1995a) proposed another algorithm with the same goals as LDB, which they referred to as discriminant pursuit (DP). DP uses the same wavelet-packet analysis as LDB. Instead of being based on relative entropy, DP is based on a simple contrast function which is determined from the means of each class (e.g., subtract mean response of one class from another). Features are extracted from the contrast function in a sequential and iterative manner. At each iteration, the portion of the signal captured by the best discriminanting basis function is then removed from the initial contrast function and the residual contrast is broken down in the wavelet domain to obtain additional features. This aspect of the method is referred to as a "pursuit" type of decomposition (see Friedman, 1987). LDB and DP achieve similar levels of performance on benchmark data sets (e.g., Breiman's waveform data) and real-world signals such as speech data (see Buckheit and Donoho, 1995a). However, experience has shown that features derived with DP are often more readily interpretable due to the fact that they are obtained from the contrast functions. In the case of a two-class discrimination problem, positive aspects of the features are related to one class of signal and negative aspects to the other.

Motivated by the work of Richmond and Optican and the novel wavelet-based methods developed by the Coifman and Donoho groups, an application of the discriminant pursuit algorithm for quantifying temporal encoding of stimuli by neuronal spike trains was developed (Laubach, 1997). The method has since been applied to many different types of neuronal data sets, including recordings of spike trains and local field potentials in the cerebral cortex and other brain areas of rats and monkeys (Nicolelis et al., 1999; Laubach et al., 2000 and unpublished data). DP identifies features on a trial-by-trial basis using single-trial peri-event histograms. The amount of information contained in neuronal firing patterns is quantified using statistical classifiers and information theory. In this manuscript, the basic discriminant pursuit algorithm is described and its application to neuronal spike train data is illustrated. In addition, a simulated data set that differs exclusively in the timing of spikes is used to directly compare the performance of DP with the PCA-based method of Richmond and Optican (1987).

2. Materials and methods

2.1. General data analysis procedures

All analysis procedures described in this study were done using custom scripts written for Matlab (The Mathworks, Natick, MA) and the Wavelab (v.8) toolbox from Stanford University http://www-stat.stanford.edu/~wavelab/. In the spirit of reproducible research (Buckheit and Donoho, 1995b), all code needed to replicate the analyses described in this manuscript can be downloaded from http://spikelab.jbpierce.org/dp.

Discriminations between trials with correct and error responses (physiological data) or response components 1 or 2 (simulations) were performed as follows (Fig. 1A): Single-trial peri-event histograms (1 ms bins) were constructed for the epoch of interest (e.g., for the physiological data, this was a 200 ms epoch). The spike trains were then smoothed over 5 ms intervals by low-pass filtering (using the `filtfilt` function in Matlab and a five-point Hamming window) and re-sampling (using Matlab's `decimate` function). In a previous study (Laubach et al., 2000), the temporal precision of firing patterns in the physiological data were examined by comparing classifications based on "raw"
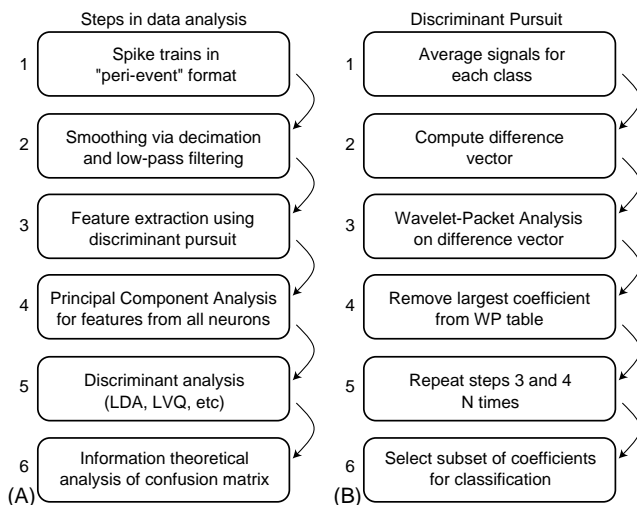
Steps in data analysis

1. Spike trains in "peri-event" format
2. Smoothing via decimation and low-pass filtering
3. Feature extraction using discriminant pursuit
4. Principal Component Analysis for features from all neurons
5. Discriminant analysis (LDA, LVQ, etc)
6. Information theoretical analysis of confusion matrix

(A)

Discriminant Pursuit

1. Average signals for each class
2. Compute difference vector
3. Wavelet-Packet Analysis on difference vector
4. Remove largest coefficient from WP table
5. Repeat steps 3 and 4 N times
6. Select subset of coefficients for classification

(B)

Fig. 1. The series of operations in the data analysis procedures described in this manuscript are shown in A. In B, the steps in the discriminant pursuit algorithm of Buckheit and Donoho (1995a) are shown. See the text for details.

spike trains (i.e., single-trial peri-event histograms with 1 ms bins) and spike trains smoothed over several different temporal epochs (i.e., 2, 5, 10, and 20 ms). Here, we used five-fold decimation as this level of smoothing was found in the study above to result in optimal levels of classification. Next, firing patterns associated with correct and error trials were identified for each neuron using discriminant pursuit (see below) and were then fed into a statistical classifier, either linear discriminant analysis or learning vector quantization (LVQ) (Kohonen, 1997; see Nicolelis et al., 1999 for details on the use of these methods for spike train data).

For LVQ, the following parameters were used for running the implementation of the method that is included in the Matlab Neural Networks toolbox: First, the learning rule from the original Matlab code was altered to perform Kohonen's optimized-learning-rate learning vector quantization (OLVQ). Second, the number of "codebook vectors" (implemented as competitive neurons in Matlab) was equal to twice the number of classes of signals. Third, the codebook vectors were initialized using a modification of the Matlab function *initlvq*. Rather than using the entire distribution of observed signals to determine the initial values and group assignments for the competitive neurons, the initial functions were assigned as the means for each class of signal. This was done to improve the performance of LVQ for data sets that have unequal numbers of trials per class (as in the neurophysiological data described in this manuscript). In such cases, the default initialization procedure from the Matlab Neural Networks toolbox tends to produce classifiers that are highly biased to the class with the largest number of observed trials in the training data.

Leave-one-out cross-validation (LOO-CV) (i.e., all trials except one used for training and a single trial used as test data) was used to estimate error rates on training and testing data and to construct confusion matrices. LOO-CV

was also used for feature extraction, although this is typically not necessary for the DP algorithm when trimmed means are used to compute the difference vector (see below). Results were expressed in terms of percentage of single trials that were classified correctly and in bits of information, defined by applying information theory (Krippendorff, 1986) to the confusion matrices. Chance levels of classification were determined using a permutation test (Efron and Tibshirani, 1994) in which the group identities were scrambled pseudo-randomly.

For feature extraction with DP, the Daubechies four-point wavelet, d4, was used for all analyses in this study. Experience has shown that this basis function is well suited for smoothed neuronal spike trains, especially for peri-event histograms that have asymmetric components (e.g., preparatory discharge) that are typical for recordings from associative and motor-related brain areas. Smoother members of the Daubechies family and other types of wavelets (Coiflets, Symmlets) typically result in lower levels of classification and produce less consistent features when multiple permutations of a given data set are evaluated. The Haar wavelet (a simple basis function that is only localized in the time domain) might also seem to be a reasonable choice for our analyses. This wavelet works very well when spikes are not smoothed prior to analysis with the DP algorithm. But for smoothed spike trains, the Haar wavelet is not well suited as it is unable to account for structure in the frequency domain.

In addition to running DP on spike trains from single neurons prior to classification, one can feed features from a collection of simultaneously recorded neurons into a classifier such as LVQ. Alternatively, once can use PCA for dimension reduction prior to classification. This can be done in two ways. First, one can arrange data from each neuron into a single column, by stacking up the set of bins for each trial end-on-end. Then, a matrix is formed from the collection of columns for all simultaneously recorded neurons. (Laubach et al. (1999) used this data structure for analysis of spike trains with independent component analysis.) This approach uses PCA (or ICA) to quantify simultaneous covariance/correlation between neurons into a reduced set of component functions. Typically, classifications with such functions perform at levels that are worse than classifications performed on the original data, due to the fact that simultaneous correlations are only a portion of the total variance in the spike trains (Laubach et al., 2000).

A second use of PCA for preprocessing is to run PCA on the collection of features identified by DP for all neurons together. This use of PCA allows for identification of a reduced set of components that reflect correlations between the wavelet features from multiple, simultaneously recorded neurons. While not the focus of the present paper, this approach has been shown to be most useful when linear discriminant analysis is used for classification and there are many neurons to be analyzed. In such a case, even with the use of DP, the data set is often still underdetermined with more predictor variables (features) than observations (trials).

2.2. The discriminant pursuit algorithm

The sequence of operations in the discriminant pursuit algorithm is described in Fig. 1B. For the sake of simplicity, the algorithm is illustrated here for a two-class discrimination. The first step in the algorithm is to find structure that contrasts neural activity in the two categories. Formally, this amounts to maximizing the difference vector, d, defined as:

$$d = \bar{x}_1 - \bar{x}_2 = \Delta + \frac{1}{\sqrt{N}}\tilde{z}, \quad \tilde{z} \sim N(0, 1) \tag{1}$$

where $\bar{x}_1$ and $\bar{x}_2$ are the average signals for each of two classes, Δ is the true distance between the classes, N is the number of samples, and $\tilde{z}$ is random noise (Buckheit and Donoho, 1995a). In practice, trimmed means (the upper and lower 10% of trials omitted from estimation of the mean) were used in the analysis described here. The difference vector is then partitioned into its fundamental time–frequency components by a discrete wavelet packet analysis (see Chapter 7 in Wickerhauser, 1994 for review).

The wavelet packet decomposition performs an iterative dyadic decomposition of the group difference vector into low and high frequency bands through convolution and decimation operations (see Fig. 4). The method thus requires that the length of the signal is of a power of two, $N = 2^J$. For example, if a signal has 64 bins, then $N = 64$ and $J = 6$. The decomposition results from operation of a low-pass filter $G = \{g_k\}_{n=0}^{L-1}$ and a high-pass filter $H = \{h_k\}_{n=0}^{L-1}$ on the signal. The filters G and H are called quadrature mirror filters, which satisfy the orthogonality constraint:

$$GH^* = HG^* = 0 \text{ and } G^*G + H^*H = I \tag{2}$$

Each decomposition of the signal results in a finer and finer partition of the frequency axis, which is assigned to an increasing depth within the wavelet packet table. The maximum degree of decimation and smoothing, which is the "maximum depth" in the wavelet packet table, is equal to J.

A wavelet packet table represents an original signal in the time–frequency (phase) plane. Each component in the wavelet packet table is known as a time–frequency atom and the collection of atoms is called a time–frequency dictionary (see Chapter 10 in Wickerhauser, 1994 for review). Each atom, Q, is indexed by its depth in the table (d), its degree of oscillation (b), and its location in time (k), i.e., $Q_{d,b,k}$. The atoms representing the signal are ordered by their amplitude, a. The atoms accounting for the largest portions of the difference vector (with the largest amplitude), $Q_{\max}$, are thus chosen as the feature that best discriminating basis.

Prior to the identification of additional features, features are extracted from the initial difference vector to give a residual difference vector, d_{res}. For each iteration, k, the residual difference vector, d_{res}^k, is equal to the previous difference vector, d_{res}^{k-1}, minus the product of the time–frequency atom, Q_k, and its associated amplitude, a_k:

$$d_{\text{res}}^k = d_{\text{res}}^{k-1} - a_k Q_k \tag{3}$$

The complete set of features is determined by performing the wavelet packet decomposition on the series of residual difference vectors.

It is possible to automatically choose a reduced set of features for classification, for example, using scree plot analysis. Here, the amplitude accounted for by each component is plotted in decreasing order and a line is fit to the upper half of the distribution. The y-intercept and slope for this line is determined and those components that have larger amplitudes than is expected by this line (i.e., that are larger than the y-intercept) are considered to be potentially useful for classification.

2.3. Neurophysiological data

Data from the published study by Laubach et al. (2000) were used to evaluate the DP method. Briefly, adult, male rats (Long Evans) were trained to perform a reaction-time task. The task required that the rats maintained a lever-press over a variable delay period and then released the lever when a trigger stimulus (tone or vibration of the lever) was presented. Correct trials consisted of the animals maintaining the lever press over the delay period and releasing the lever with a reaction time less than 1 s. Error trials occurred when the lever was released before the trigger stimulus. Chronic and simultaneous recordings of neuronal spike trains were obtained in the forelimb area of the rat motor cortex as the animals learned the task and in fully trained subjects. Procedures for chronically implanting arrays of microwire electrodes (NB Labs, Dennison, TX) and data acquisition, using a multielectrode single-unit recording system (Plexon, Dallas, TX), are described elsewhere (Nicolelis et al., 1997a). All rats were treated in accordance with NIH guidelines.

In the data set used for this study, 19 neurons were recorded simultaneously in layer V of the rat motor cortex (in the medial and lateral agranular cortex). The rat had been trained to perform the reaction-time task for more than two months. There were 133 correct and 55 error trials in the 1-h session. Thus, the chance level of classification was 70.7%.

2.4. Simulated spike trains

A series of point events from an actual behavior experiment were used to create simulated spike trains with realistic statistical and qualitative features to those obtained in actual neuronal ensemble recording experiments. Baseline signals, used in all simulations, were 2 Hz random processes with inter-event intervals selected from the Poisson distribution and that occurred over the entire epoch from 406 ms prior to the node (time 0) to 106 ms after the node. The first event after the occurrence of the trigger stimulus was used as an onset for the epoch containing the signal components (Fig. 2A) that was added to the baseline activity. These signals were analogous to what one would expect to find if a hidden event in the nervous system occurred during behavior and elicited a
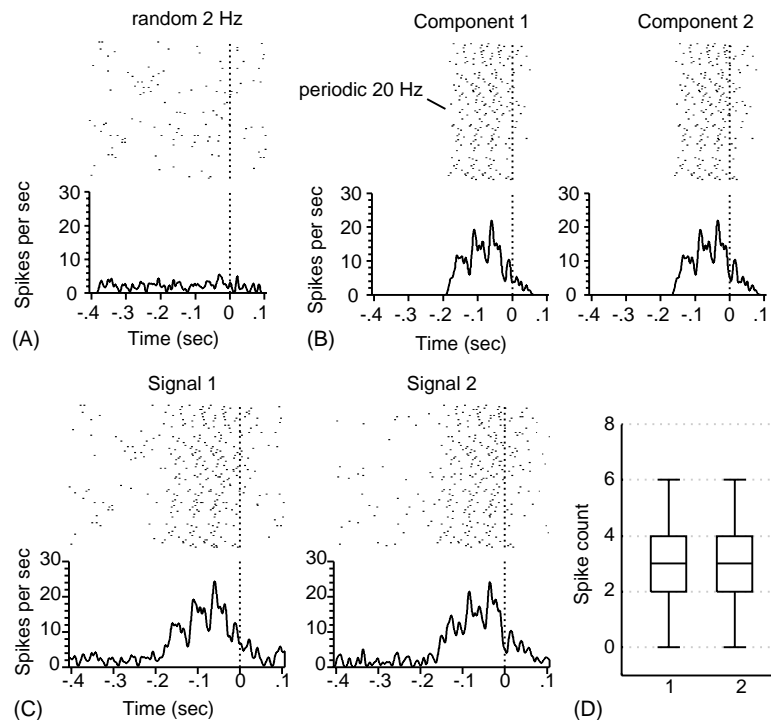
Fig. 2. The construction of the simulated spike trains used in this study is illustrated. In A, the random background discharge is shown in the raster (upper) and peri-event histogram (lower) plots. In B, the two periodic components, based on a triplet of spikes, are shown. These two signals were combined with the background activity in A to give the two classes of simulated signals shown in C. Note that while the signals differed in the timing of the triplet of spikes there was no difference in the total spike count (D).

neuronal response that contained some degree of information about the animal's performance of the behavioral task. The signal components were periodic processes that consisted of three events, with 50 ms intervals, that occurred after the first event after the trigger stimulus (Fig. 2B). The events in one of the periodic components were simply shifted by 25 ms to give the other periodic component. The signal components were then combined with the baseline activity to give the signals used in the analyses (Fig. 2C). Both signals contained an increase in the time-averaged firing rate for the epoch 200 ms prior to the node (time 0), with time-averaged rates of 11.85 ± 1.50 Hz and 2.37 ± 0.30 spikes. There was no difference therefore in the time-averaged firing rates for the signals (Fig. 2D): box plots of the total number of spikes in the 200 ms epoch prior to the node showed that the signals contained equivalent numbers of spikes (median = three spikes). The goal of the analysis was to determine if any available method for spike train analysis could detect the phase difference between the two classes of signals.

3. Results

3.1. Neurophysiological data

The DP method is illustrated for a neuron recorded in the rat motor cortex during the performance of a simple reac-

tion time task (Laubach et al., 2000). In this task, trials with correct responses (i.e., a motor response was sustained over a delay period until a trigger stimulus was presented) were compared with trials with error responses (i.e., the rat responded prior to the scheduled time of the trigger stimulus). The goal was to determine if neuronal activity in rat motor cortex was different on these types of trials. A difference in activity might reflect the encoding of the trigger stimulus by the motor cortex or of the animal's ability to inhibit responding on the error trials (see Laubach et al., 2000 for discussion of this task).

Prior to application of DP, the spike trains for each trial were smoothed using low-pass filtering and five-fold decimation (see Section 2). This process is illustrated in Fig. 3A, where the smoothed spikes can be seen as an approximation of the time-varying firing rate or spike probability, which was referred to as the spike density by Richmond and Optican (1987). Average signals for correct and error trials are shown in Fig. 3B. The neuron fired at an increasing rate over the final 100 ms of the foreperiod on the correct trials (solid black line) and was otherwise not modulated on the error trials (dashed gray line). The difference vector (Fig. 3C) revealed a trend for the neuron to fire at a higher rate immediately prior to movement onset on the correct trials.

Decomposition of the difference vector using DP resulted in the distribution of amplitudes for the complete set of basis functions shown in Fig. 3D. The first six basis functions (Fig. 3E) accounted for larger portions of the difference
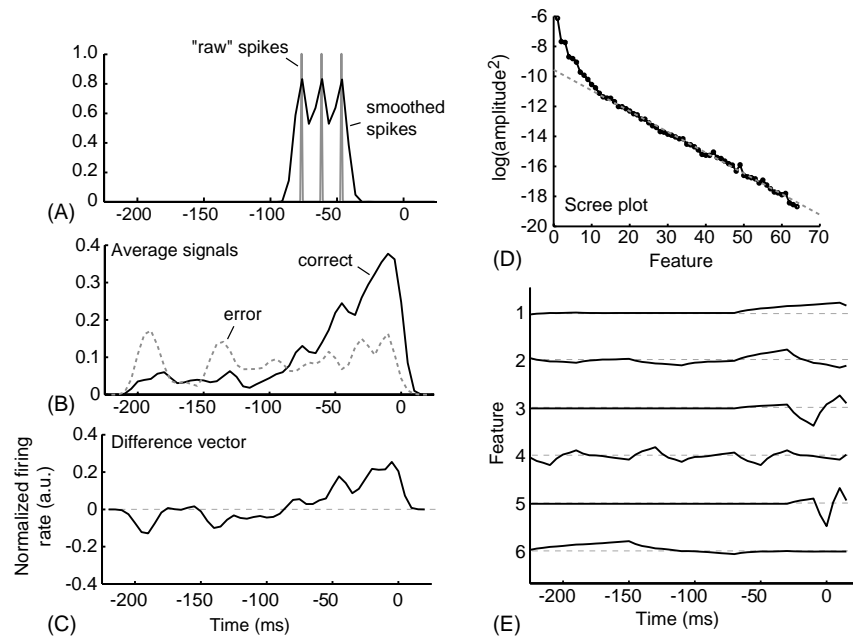
Fig. 3. Application of the discriminant pursuit method to a spike train recorded from a neuron in the rat motor cortex during the performance of a simple reaction time task (Laubach et al., 2000) is shown. In A, the smoothing procedure for converting the impulse train into a smooth representation of spike density is illustrated. In B, the mean signals for each type of trial, correct and error responses (see the text for details) are shown. In C, the difference between the two mean signals is illustrated as the so-called "difference vector". This signal was decomposed using the discriminant pursuit method. The amplitudes of the series of coefficients identified with the method are shown in the scree plot in D. A reduced set of six features with the largest amplitudes was used for subsequent analyses. In E, the time course of the identified features, or firing patterns, is shown.

vectors than was anticipated based on the distribution of the amplitudes of the complete set of basis functions. The first basis function was localized to the final 100 ms of the trial, i.e., the portion of time when the mean signals for the correct and error trials were most different for this neuron. The second basis function reflected the slight increase in firing rate that was apparent in the mean response for the correct trials at approximately 50 ms prior to movement onset. The third and fifth basis functions were based on the time of movement onset, which was the time with the largest difference between the two types of signals. The fourth function was not well localized and accounted for the overall shape of the difference vector, especially around times earlier in the histogram when the neuron fired at slightly higher rats on the error trials. Finally, function 6, which was near the level of "noise" expected from the scree analysis above, further emphasized differences in firing rates between the correct and error trials around 150 ms prior to movement onset.

Dyadic tiling of the time–frequency domain by wavelet-packet decomposition is illustrated schematically in Fig. 4A. An actual wavelet packet table for the difference vector, or "contrast", obtained from the data in Fig. 3 is shown in Fig. 4B. The initial difference vector, with a length of 64 samples, is shown in the upper row (depth 0). The second row stores the result of applying a single decimation to the signal and stores the smooth (low-pass) component on the left side ($s1$–32) and the detail (high-pass) component on the right side ($d1$–32). A similar splitting of the signal is then
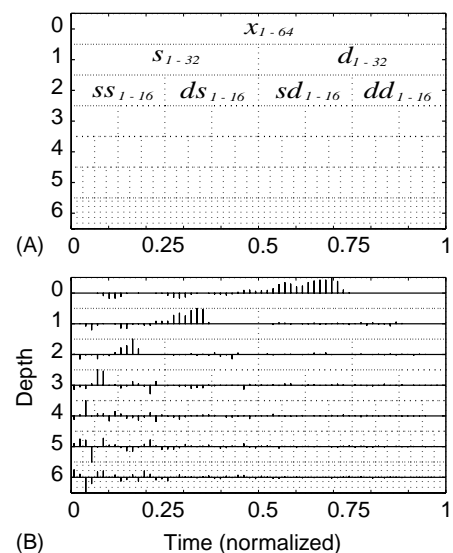

Fig. 4. Wavelet packet decomposition is used in the discriminant pursuit method to decompose a signal into its essential components that are well localized in the time and frequency domains. In A, the standard wavelet packet table is shown for a signal of length 64 and a maximum depth of decomposition of 6. See the text for details on this method for signal decomposition. In B, the wavelet packet table for the difference vector from Fig. 3 is shown. The energies of the signal components are depicted in the line plots at each depth in the wavelet packet table. Note that almost all of the signal energy is concentrated in the smooth aspects of the signal (i.e., on the left side of the table).
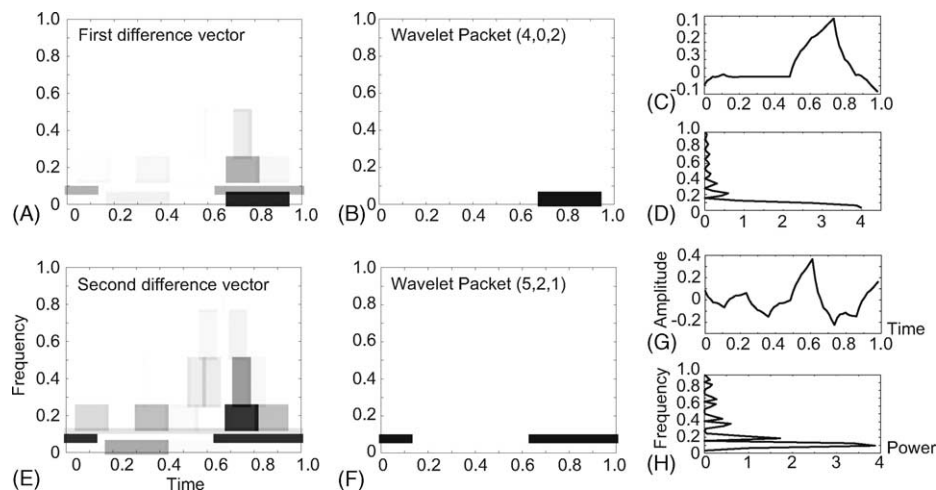
Fig. 5. Partitioning of the difference vector in the time and frequency domains using the discriminant pursuit method. In A, the wavelet components for the difference vector from the neuron shown in Fig. 3 are depicted in the time–frequency plane. The components have been weighted using the best basis algorithm for the purposes of illustration. The signal component that accounts for the largest portion of the difference vector is removed and the difference vector is re-expressed. The initial basis function is called the "best discriminating basis" and is shown in the time–frequency plot in B and in the time and frequency domains in C and D, respectively. The feature was localized in the wavelet packet table at a depth of 4, a degree of oscillation of 0, and a temporal location of 2. After the first discriminating basis function is removed from the difference vector, the residual difference vector (E) is then decomposed with wavelet packet decomposition. The next largest component of the signal is shown in F–H. This is the second best discriminating basis and it was localized in the wavelet packet table at a depth of 5, a degree of oscillation of 2, and a temporal location of 1.

done for each of the representations at depth 1 resulting in a smooth-smooth component, a detail-smooth component, and smooth-detail component, and a detail-detail component, each with length 16. This process is further repeated to a depth of 6.

The sequential breakdown of the difference vector into a series of features can also be visualized in the time and frequency domains (Fig. 5). A time–frequency representation of the initial difference vector is shown in Fig. 5A. The contributions of each wavelet component have been weighted via the "best basis" algorithm (Coifman and Wickerhauser, 1992) to emphasize the amount of variance in the difference vector that was accounted for by each component. In Fig. 5B, the best discriminating basis function is shown in the time–frequency plane. In the time domain (Fig. 5C), this function was localized to the second half of the time prior to movement onset. In the frequency domain (Fig. 5D), this function was mostly localized to the low frequency range (~0.1 times the Nyquist frequency or 10 Hz for spikes sampled at a 5 ms precision). The residual difference vector is shown in Fig. 5E and the location of the next best discriminating basis is in Fig. 5F. This function exhibited a larger degree of fluctuation during the time prior to movement (Fig. 5G) and this was reflected in the frequency domain (Fig. 5H), where there were multiple frequencies around 10 and 20 Hz accounted for by the basis function.

Classifications based on smoothed spike trains (five-fold decimation) and the LDA and LVQ methods were correct for 56.4 and 71.8% of trials (0.003 and 0.147 bits), respectively. The improved performance of LVQ over LDA is likely due to the large number of bins in the spike trains that results in having an underdetermined data set. When DP-based fea-

tures for this neuron were fed into the LDA and LVQ classifiers, 75.5 and 70.2%, respectively, of trials were classified correctly as correct or error responses (0.076 and 0.134 bits of information). Despite the percentage correct being higher for LDA than LVQ, LDA produced less well-balanced confusion matrices and thus provided less total information about the animal's performance of the task. In fact, the LDA classifier was highly biased to the correct response trials and correctly predicted the class labels for 123 of 133 trials with correct responses (92.5% correct) and only 19 of 55 trials with error responses (34.5% correct). By contrast, LVQ was biased toward the error response trials and correctly predicted the class labels for 88 correct responses trials (66% correct) and 44 error response trials (80% correct). This result shows why it is essential to check the confusion matrix whenever a multivariate method is used for discriminant analysis as the percentage correct metric if used alone can be misleading (see the on-line supplementary material from Laubach et al., 2000).

At the ensemble level (i.e., for all 19 neurons that were recorded simultaneously), the following results were obtained using five-fold decimation and low-pass filtering on the spike trains: LDA_{raw}—45.7% correct, 0.003 bits; LDA_{DP}—70.7% correct, 0 bits (all trials classified as correct responses); $LDA_{DP \cdot PCA}$—79.3% correct, 0.151 bits; LVQ_{raw}—73.9% correct, 0.096 bits; LVQ_{DP}—79.3% correct, 0.202 bits; $LVQ_{DP \cdot PCA}$—77.7% correct, 0.177 bits. Here, *raw* is for the direct application of LDA and LVQ to the spike trains, DP is for using DP processing prior to LDA or LVQ, and DP·PCA is for following DP with PCA on the collection of features from all neurons. The best overall results were obtained with the LVQ_{DP} method and

this analysis provided ∼14% more information than the next best classification scheme (LVQ$_{DP.PCA}$). As in the case of the direct application of LDA, the LDA$_{DP.PCA}$ classifier was biased towards the correct responses. Finally, analyses without DP prior to the combination of PCA and LDA or LVQ produced the following results: LDA: 69.2% correct, 0.024 bits; LVQ: 62.7% correct, 0.032 bits.

3.2. Simulated spike trains

Feature extraction using DP and PCA (Richmond and Optican, 1987) was compared for simulated spike data that differed only in the precise timing of a triplet of spikes (see Section 2). Eigenvectors for the first three principle components are shown in Fig. 6A. The first component was similar to the overall average response for signal components 1 and 2. Power spectral analysis showed that the higher components accounted for the frequency content of the triplet of spikes, with maximum power in the range of 20 and 40 Hz (Fig. 6B). The 40-Hz component accounted for the difference in the phase of the signals. Classifications of

the PCA-based features with LVQ were correct for 61.3% of trials (0.037 bits of information with chance equal to 50%). Features extracted with DP improved on the results obtained with PCA. Average signals for signal class and the difference vector are shown in Fig. 6C–D. The best discriminating basis function was periodic and accounted for the frequency range of the true signal difference with maximum power in the frequency domain at 20 Hz (Fig. 6E–F). The second and third features accounted for the timing of the onset and offset, respectively, of the periodic signals. These features allowed for the classification of the signals by their phase (64.5% correct, 0.065 bits, 1.75 times that provided by the PCA-based features). Also, classifications by both methods were significantly better than expected by chance ($P = 0$; 1000 repetitions with a permutation test, i.e., class labels randomly permuted). Chance levels of prediction for this data set averaged around 48.6% correct (STD = 2.8) and provided 0.003 bits of information (STD = 0.003).

4. Discussion

Spike train data from the rat motor cortex and simulated data sets were used to show that discriminant pursuit (DP) outperforms principal component analysis (PCA) in accounting for temporally localized events in spike trains that vary as a function of experimental condition (e.g., across a range of sensory stimuli or different types of motor acts). A major strength of the method is that the dimensionality of the spike train data can be greatly reduced following extraction of the relevant features. As a result, methods for discriminant analysis, such as Linear Discriminant Analysis or methods based on artificial neural networks (e.g., LVQ), are able to operate on a much more efficient representation of the data and can therefore provide significantly better statistical predictions than if the raw spike trains were directly analyzed with a statistical classifier.

In the studies described in this manuscript, we used five-fold decimation and low-pass filtering to convert "raw" spike trains (i.e., single-trial peri-event histograms with 1 ms bins) into continuous signals that estimated the density of spikes over time. This value of decimation was chosen based on prior work (Laubach et al., 2000) in which a range of bin sizes was compared: The outcomes of single trials were predicted equally well with five-fold (79.6 ± 1.9% correct, 0.207 ± 0.032 bits) and 10-fold (77.3 ± 1.4%, 0.178 ± 0.025 bits) decimation. However, predictions were significantly worse (ANOVA: $P < 0.005$) with 20-fold decimation (71.8 ± 0.7% correct, 0.105 ± 0.008 bits) or when a single bin of 200 ms was used for classification (63.6 ± 1.9% correct, 0.038 ± 0.008 bits). We also observed that there is no improvement over the results above when time scales finer than 5 ms are used as input to DP and LVQ and that application of LVQ to "raw" peri-event histograms (i.e., 1 ms bins) results in classifications at chance levels. The time scale of neuronal activity in the rat motor cortex
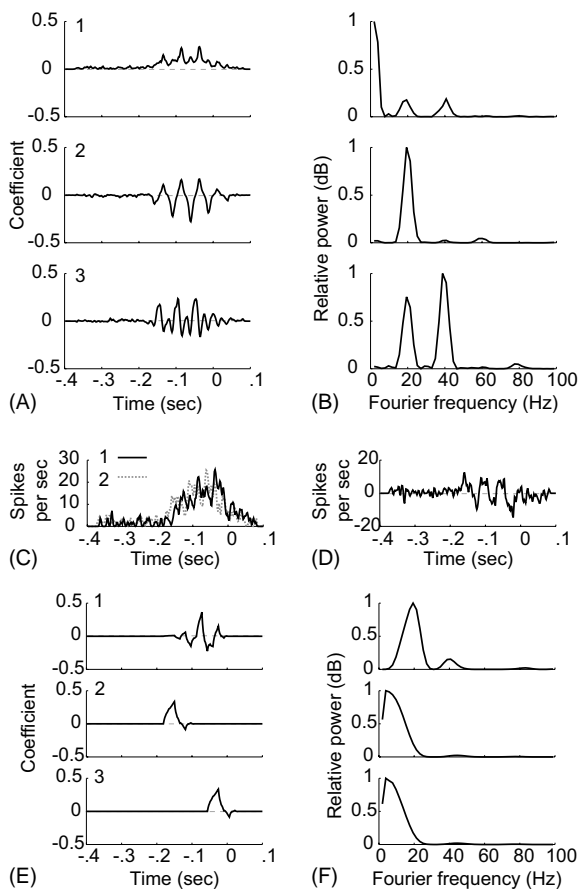

Fig. 6. The decompositions of the simulated spike trains are shown for the PCA (A–B) and DP (C–F) preprocessing methods. The top three features identified with PCA are shown in the time (A) and frequency (B) domains. In C and D, the means signals and difference vector for the two classes of signals are shown. This signal was decomposed into the set of features using the DP method (E—time domain, F—frequency domain).

during reaction time performance therefore seems to be on a short time scale (<10 ms). Resolution of information processing on this time scale seems to require preprocessing with methods like DP. However, this may not be the case in other neuronal systems or other types of behavioral tasks or sensory mapping experiments. It is thus important to examine this issue by comparing spike trains at different temporal resolutions in future studies that use methods such as DP.

Despite the strengths of the DP method described above, the algorithm is not flawless. Intrator and colleagues (1997) pointed out that DP is not able to discriminate between certain combinations of purely periodic functions. This is due to potential cancellations between such functions in the process of computing the difference vector. However, this is not a practical issue for spike train data, as pure sinusoidal oscillations are rarely seen. In such a case, the local discriminant basis method of Saito and Coifman (1994), which is not affected by this issue, could serve as an alternative to DP. Intrator et al. (1997) also commented that in their experience DP performs worse than LDB when applied to high-dimensional data that are measured over a limited number of trials. However, our experience using these two methods has not revealed such a difference in performance. In fact, we have repeatedly observed that for spike data and a given wavelet (e.g., Haar or d4) the two methods typically converge to a highly similar reduced set of features.

The use of a method such as DP for preprocessing spike trains can help resolve a major issue in systems neuroscience: the extent to which neurons make use of temporal or rate based codes for transmitting information (Engel et al., 1992; Shadlen and Newsome, 1995; Softky, 1995; Theunissen and Miller, 1995). The DP method described here is well-suited to this issue due to its ability to reveal changes in overall firing rate (which results in a single flat wavelet function; Laubach, unpublished observations) and to localize fine temporal patterns in spike trains that vary as a function of the parameters of experimental stimuli, movements, or the performance of a behavioral task.

The DP method has much potential for wider application within neurophysiology (e.g., spike sorting and analysis of local field potentials, EEG, and EMG). The method is relatively fast and inherently parallel, allowing it to be used for studies of real-time decoding of brain signals. Indeed, we have recently implemented an on-line version of the method for performing real-time analyses of neurophysiological data (Laubach, 2003). In this context, the method may provide a means for experimental validations of putative brain codes during neurophysiological recording sessions in awake, behaving subjects. The DP method is also applicable to non-spiking data such as EMG data (Laubach et al., 2000) and EEG signals collected in the context of brain-computer interfaces (Wolpaw et al., 1991). Indeed, our initial application of the method to such signals has shown that it is a strong candidate for such signals (see the results from the BCI2003 data analysis competition).

References

Buckheit J, Donoho DL. Improved linear discrimination using time–frequency dictionaries. Proc SPIE 1995a;2569:540–51.

Buckheit J, Donoho DL. Wavelab and reproducible research. Wavelets and Statistics. A. Antoniadis. New York: Springer-Verlag; 1995b.

Coifman RR, Wickerhauser MV. Entropy based algorithms for best basis selection. IEEE Trans Inform Theory 1992;38:712–8.

Deadwyler SA, Bunn T, et al. Hippocampal ensemble activity during spatial delayed-nonmatch-to-sample performance in rats. J Neurosci 1996;16:354–72.

Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman and Hall; 1994.

Engel AK, Konig P, et al. Temporal coding in the visual cortex: new vistas on integration in the nervous system. Trends Neurosci 1992;15:218–26.

Fisher RA. The use of multiple measurements in taxonomic problems. Ann Eugenics 1936;7:179–88.

Friedman JH. Exploratory projection pursuit. J Am Stat Assoc 1987;82:249–66.

Furukawa S, Middlebrooks JC. Cortical representation of auditory space: information-bearing features of spike patterns. J Neurophysiol 2002;87:1749–62.

Ghazanfar AA, Stambaugh CR, et al. Encoding of tactile stimulus location by somatosensory thalamocortical ensembles. J Neurosci 2000;20:3761–75.

Gochin PM, Colombo M, et al. Neural ensemble coding in inferior temporal cortex. J Neurophysiol 1994;71:2325–37.

Hastie T, Tibshirani R, et al. The elements of statistical learning. New York: Springer; 2001.

Heller J, Hertz JA, Kjaer TW, Richmond BJ. Information flow and temporal coding in primate pattern vision. J Comput Neurosci 1995;2:175–93.

Hyvarinen A, Karhunen J, et al. Independent component analysis. New York: Wiley; 2001.

Intrator N, Huynh QQ, et al. Wavelet feature extraction for discrimination tasks. In: Proceedings of the 1997 Canadian Workshop on Information Theory, 1997.

Kohonen T. Self-organizing maps. New York: Springer-Verlag; 1997.

Krippendorff K. Information theory: structural models for qualitative data. Thousand Oaks, CA: Sage Publications; 1986.

Laubach M. A neurophysiological investigation of the role of the cortico-basal ganglia system in reaction-time performance. Unpublished doctoral dissertation, Neuroscience Program. Winston-Salem, NC: Wake Forest University; 1997.

Laubach M, Arieh Y, et al. A cluster of workstations for on-line analyses of neurophysiological data. In: Proceedings of the 29th Annual Northeast Bioengineering Conference; 2003.

Laubach M, Shuler M, et al. Independent component analyses for quantifying neuronal ensemble interactions. J Neurosci Methods 1999;94:141–54.

Laubach M, Wessberg J, et al. Cortical ensemble activity increasingly predicts behaviour outcomes during learning of a motor task. Nature 2000;405:567–71.

Miller EK, Li L, et al. A neural mechanism for working and recognition memory in inferior temporal cortex. Science 1991;254:1377–9.

Nicolelis MA, Ghazanfar AA, et al. Reconstructing the engram: simultaneous, multisite, many single neuron recordings. Neuron 1997a;18:529–37.

Nicolelis MA, Lin RC, et al. Neonatal whisker removal reduces the discrimination of tactile stimuli by thalamic ensembles in adult rats. J Neurophysiol 1997b;78:1691–706.

Nicolelis MA, Ghazanfar AA, et al. Simultaneous encoding of tactile information by three primate cortical areas. Nat Neurosci 1998;1:621–30.

Nicolelis MA, Stambaugh CR, et al. Methods for simultaneous multisite neural ensemble recordings in behaving primates. In: Nicolelis MA, editor. Methods for neural ensemble recordings. Boca Raton: CRC Press; 1999. p. 121–56.

Penev PS, Atick JJ. Local feature analysis: a general statistical theory for object representation. Network: Comput Neural Syst 1996;7:477–500.

Richmond BJ, Optican LM. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform. J Neurophysiol 1987;57:147–61.

Ripley BD. Pattern recognition and neural networks. Cambridge: Cambridge University Press; 1996.

Saito N, Coifman RR. Local discriminant bases. Proc SPIE 1994;2303:2–14.

Schoenbaum G, Eichenbaum H. Information coding in the rodent prefrontal cortex. II. Ensemble activity in orbitofrontal cortex. J Neurophysiol 1995;74:751–62.

Shadlen MN, Newsome WT. Is there a signal in the noise? Curr Opin Neurobiol 1995;5:248–50.

Softky WR. Simple codes versus efficient codes. Curr Opin Neurobiol 1995;5:239–47.

Theunissen F, Miller JP. Temporal encoding in nervous systems: a rigorous definition. J Comput Neurosci 1995;2:149–62.

Vapnik VN. The nature of statistical learning theory. New York: Springer; 2000.

Wickerhauser MV. In: Natick MA, Peters AK, editors. Adapted wavelet analysis from theory to software; 1994.

Wolpaw JR, McFarland DJ, et al. An EEG-based brain-computer interface for cursor control. Electroencephalogr Clin Neurophysiol 1991;78:252–9.