Joint probability-based neuronal spike train classification

Yan Chen, Vitaliy Marchenko and Robert F. Rogers*

Department of Electrical and Computer Engineering, University of Delaware, Newark, DE, USA

(Received 1 February 2008; final version received 18 August 2008)

Neuronal spike trains are used by the nervous system to encode and transmit information. Euclidean distance-based methods (EDBMs) have been applied to quantify the similarity between temporally-discretized spike trains and model responses. In this study, using the same discretization procedure, we developed and applied a joint probability-based method (JPBM) to classify individual spike trains of slowly adapting pulmonary stretch receptors (SARs). The activity of individual SARs was recorded in anaesthetized, paralysed adult male rabbits, which were artificially-ventilated at constant rate and one of three different volumes. Two-thirds of the responses to the 600 stimuli presented at each volume were used to construct three response models (one for each stimulus volume) consisting of a series of time bins, each with spike probabilities. The remaining one-third of the responses were used as test responses to be classified into one of the three model responses. This was done by computing the joint probability of observing the same series of events (spikes or no spikes, dictated by the test response) in a given model and determining which probability of the three was highest. The JPBM generally produced better classification accuracy than the EDBM, and both performed well above chance. Both methods were similarly affected by variations in discretization parameters, response epoch duration, and two different response alignment strategies. Increasing bin widths increased classification accuracy, which also improved with increased observation time, but primarily during periods of increasing lung inflation. Thus, the JPBM is a simple and effective method performing spike train classification.

Keywords: spike train analysis; slowly adapting pulmonary stretch receptors; classification problem; temporally discretized binning procedure

1. Introduction

Neural spike trains are the internal language used by the nervous system to transmit information. Spike patterns may carry information-based on the timing of action potentials, or the sequence of interspike intervals (ISIs), only if these patterns repeat more often than they would by chance given the same average firing rate or ISI distribution [12]. Many methods have been proposed to analyse spike patterns. For example, predefined templates with small variance tolerances were used to discover favoured patterns in neuronal firing activity [4], and are useful in evaluating the statistical significance of a pattern’s temporal accuracy [15]. Zipser and collaborators [19] used patterns of firing rate rather than patterns of spike times to solve the difficulty posed by spike time tolerances. Some studies modelled the spike train by convoluting spike trains with Gaussian kernel [5,14] or exponential kernel [16], and produced relatively simple, closed equations describing neuronal activity patterns imparted by the features of the Gaussian or exponential functions. Another approach used to quantify
spike train similarity, cost-function distance [17,18], ascribes a cost to elementary transform operations of deleting, inserting and time-shifting spikes, thereby assigning a similarity metric (distance) between any two spike trains. Other strategies associated with analysing the temporal patterns involve evaluating local similarity in short segments of spike trains [1] and the ISI pattern [11]. These are based on the spike series represented by precise times, while other analyses use a coarser binning procedure. Quantifying spike trains into bins and comparing the Euclidean distance among spike trains in \( n \)-dimensional vector space has also been used to resolve the temporal structure of responses [7–9]. This general procedure is also used to create peri-stimulus time histograms, another means by which to average and compare spike train responses.

In addition to the above methods, classification has also been performed by artificial neural networks (ANNs) and statistical methods such as linear discriminant analysis (LDA), linear vector quantization (LVQ), principal component analysis (PCA) and independent component analysis (see Ref. [6] for review). Foffani and Moxon [6] concluded that the PSTH-generated, Euclidean distance-based method (EDBM) is an efficient alternative to LDA and ANNs when studying ensembles of spike train responses to discrete sensory stimuli. In a recent study, we analysed the performance of response classification using a ‘sparse’ PSTH-based response representation and found that it did not significantly reduce performance while offering marked computational savings [3]. Osan and colleagues [10] compared the effectiveness of different statistical methods in pattern classification, including multiple discriminant analysis (MDA), PCA, ANN and multivariate Gaussian distributions (MGD). They concluded that typical performance ranking of these methods on under-sampled neural data of large dimension is MDA > PCA > ANN > MGD.

In this study we developed and applied a joint probability-based method (JPBM) of classifying individual slowly adapting pulmonary stretch receptor (SAR) spike train responses evoked by three lung inflation stimuli that differed only in amplitude. Applying the same temporal binning procedure, we compared classification performance between our JPBM and the conventional EDBM. Neither EDBM nor JPBM make \textit{a priori} assumptions regarding the nature of the responses. Unlike other methods (e.g. ‘cost function’-based methods), neither the JPBM nor the EDBM use subjective weighting to determine similarity, thereby avoiding user-determined biases. The JPBM was constructed by calculating the probability of observing many consecutive individual spike/non-spike events, in which a ‘test’ spike train to be classified serves as a probability filter to select the best (i.e. the one with the maximal joint probability) of three model patterns generated in response to three different stimuli. The Euclidean distance between an individual spike train and the same three model responses were calculated, and the classification accuracy for both methods was compared after systematic variation in analysis parameters, including discretization bin width, response alignment method, and duration of the response/model used. This allowed for direct comparison of the JPBM and EDBM over a range of parameter values.

2. Materials and methods

Data were collected \textit{in vivo} from five adult male New Zealand White (1.9–3.0 kg) rabbits. Anaesthesia and surgical preparation procedures were identical to those described in Refs. [3,13]. All procedures were approved by the University of Delaware’s Institutional Animal Care and Use Committee, and conformed to the standards set in the Animal Welfare Act.
2.1 Stimulus and recording techniques

During mechanical ventilation (Harvard Bioscience, Holliston, MA, USA), tracheal inflation pressure (TP) was recorded at the inspiratory sidearm of the Y-shaped tracheal tube and used as a measure of lung distension. The electrical activity of individual SARs was recorded extracellularly within the nodose ganglion using tungsten electrodes (Z ~ 1 MΩ at 1 kHz; FHC, Inc., Bowdoinham, ME, USA). Electrodes were inserted into the nodose ganglion manually without removal of the connective tissue capsule. The lead wire (between the electrode shaft to the pin connecting it to the headstage) was coiled into a spring to obtain a ‘floating’ electrode configuration [2,13]. A micromanipulator mounted on a stereotaxic frame was used to make fine adjustment of the electrode tip position if necessary. SARs were identified by their faithful responses to manipulations in TP and categorized as either low or high threshold, depending on the presence or absence of tonic activity (i.e. between lung inflations). Continuous recordings of individual SARs were maintained for >1 h, during which inflation volume was set to the three different values (9, 12 or 15 ml). Each volume was presented for >20 min, and data collection commenced after a 5 min buffer time following a switch to a new volume. TP and extracellular unit activity were collected at 10,000 samples/s and stored on the hard drive of a Pentium-based PC via a data acquisition system (PowerLab 16 and Chart® version 5.3, ADInstruments, Colorado Springs, CO, USA).

Since relatively large (> 4:1) SNRs were achieved, spike detection (0.1 ms resolution) was performed with the use of conventional software (Spike2 v. 4.22, Cambridge Electronic Design Ltd, Cambridge, UK) using a simple threshold function. Pattern analysis and classification computations were performed by custom written M-files for MATLAB® (version 6.5, The MathWorks Inc., Natick, MA, USA). Some analysis was performed using Microsoft® Excel (version 10, the Microsoft Corp.). Figures were plotted with Origin® (version 7.5, The OriginLab Corp., Northampton, MA, USA). All were executed on a conventional Pentium 4 PC (Windows® XP professional).

2.2 Spike alignment

Two different strategies were used to align spike trains: stimulus-based and response-based. The former is based on the TP stimulus waveform. Once the inflations are aligned, their corresponding spike trains are aligned by virtue of their fixed relationship to the TP waveform. Since SARs fire in response to periodic lung inflations, it is reasonable to assume that the same SAR begins firing at approximately the same TP threshold in response to the same stimulus. Thus, the second method was used to align responses to the first spike in the series, as done in previous studies [2,3].

2.3 Spike classification methods

2.3.1 Joint probability-based classification

Individual SAR responses were transformed into binary sequences by discretizing time into uniform bins and assigning values of 1 or 0 to bins that either did, or did not, contain a spike, respectively. Ventilation volume was selected as the stimulus variable. For each stimulus condition (9, 12 or 15 ml), 600 lung inflations were presented (15 min × 40 breaths/min, 50% duty cycle). Of these, two-thirds were randomly chosen and used to construct a canonical model ‘response’ of the SAR to a given stimulus, while the remaining one-third were used as responses to be classified (hereafter referred to as ‘test responses’). A model response was constructed by calculating the probability of observing a spike in each bin starting from just after the onset of the breath through the end of the breath
(see Performance evaluation, below). This sequence of probabilities defines the cell’s average firing response pattern for a given stimulus. In order to perform classification, an average response pattern (model) was created for each of the three stimuli.

Consider an individual spike train digitized into a binary representation with the same number of bins as the model responses. If it was produced in response to one of the three stimuli, it should be more similar to the model response for that stimulus than either of the other two. In order to ascertain which model was most similar to a given test response, individual test responses (of which 200 were generated from each stimulus) consisting of binary representations were used to ‘filter’ the three model responses as follows. The probability, \( P(E_k) \), of observing a particular event, \( E \), in bin \( k \) was read out for each bin in the model responses. The event, either the presence or absence of a spike in the test response, was used to determine which probability to use from bin \( k \) in a given model. For example, if the \( k^{th} \) bin in the test response did not contain a spike, and the probability of observing a spike in the same bin of the model response in question was 0.472, then the probability of observing the same event (i.e. no spike) in that model would be 0.528, and that value would be used for \( P(E_{k,M}) \) in Equation (1). In the cases where the probability of observing a spike was 0 in a particular bin, a nominal value of 0.0005 was used to avoid zero multiplication (see below). Bins are assumed to be independent based on the fact that ISIs less than axonal refractory periods were never observed. Therefore the joint probability of observing the same event in a test and model response, defined by the product of the probabilities over all bins, provides a measure of likelihood that the test response and model were evoked by the same stimulus, as long as they are composed of the same number of bins and aligned accordingly. This probability is described by the following expression:

\[
P(T|M_i) \propto \prod_{k=1}^{n} P(E_{k,M_i}),
\]

where \( T \) denotes the test spike train response, \( M_i \) denotes one of the three model responses, \( k \) represents the specific bin, and \( n \) is the number of bins used to perform the calculation. Three products (i.e. ‘joint’ probabilities, referring to probability of observing identical model and test response events over all bins) were generated with respect to a given test response, one associated with each model response. The largest product of the three indicates the maximal likelihood that the response was evoked by that stimulus (i.e. that the test response belongs to that response class).

### 2.3.2 Euclidean distance-based classification

In order to evaluate the JPBM, we also classified test responses using the EDBM. The same model responses were used as those in the JPBM, and the specific test responses were the same as those used in the JPBM. Each model response can be viewed as a point in an \( n \)-dimension space, where \( n \) is the number of bins, and each bin contains the spike probability over all 400 responses used to generate it. A test response was defined by a binary string of the same length in which 0 indicated no spike and 1 indicated a spike. Test responses were placed into the \( n \)-dimensional space, and compared directly to each of the three model responses by calculating the Euclidean distance, \( d \), between the test and three model responses. The Euclidean distance is defined as:

\[
d(M - T) = \|M - T\| = \sqrt{\sum_{i=1}^{n} (M_i - T_i)^2}.
\]

For each individual test spike train, \( T \), and model response, \( M \), the Euclidean distance, \( d(M - T) \), was computed in \( n \) dimensions. The spike train was simply classified into the
specific stimulus category to which it is closest in vector space. Notice that although models are defined by probabilities and therefore occupy intermediate values in \( n \)-space, individual test responses must occupy vertexes in this same \( n \)-space because of the digital contents of their bins. Figure 1 illustrates both classification methods.

### 2.4 Analysis parameters

Two specific analysis parameters were varied systematically: discretization bin width and observation time (duration). The bin width was varied between 1 ms and just less than the minimum ISI recorded for a given SAR, in increments of 1 ms. The minimum ISI (i.e. highest instantaneous firing rate) was invariably recorded during inflations with 15 ml. When displaying results from multiple SARs in one plot, all bin widths are normalized by their minimum ISIs. The observation time (number of bins \( \times \) bin width) was varied in increments of 5 ms or 1 bin, whichever was larger, spanning some fraction of the entire response, beginning just after the first spike (response-aligned method) or just after the onset of lung inflation (stimulus-aligned method). Both alignment strategies were used in both classification methods to compare their effect on classification accuracy.

### 2.5 Performance evaluation

One specific performance metric was chosen to evaluate the performance of the JPBM and the EDBM: maximum accuracy. Maximum accuracy denotes the fraction of correct assignments of 600 test responses (200 drawn from each of the three stimuli) to the stimuli that evoked them. The ‘maximum’ qualifier indicates that the maximum value, over all observation times examined, was reported for a given bin width and stimulus condition. This observation time may have varied from one bin width to the next for the same SAR. Thus, the maximum accuracy identifies the best classification performance that can be achieved for a given SAR using a specific bin width and alignment strategy and within a single stimulus condition (9, 12 or 15 ml). Maximum accuracy values are provided for test spike trains evoked by separate stimulus values within an SAR, as well as for total performance in that neuron over all stimuli. This latter value, the ‘overall maximum classification accuracy’ is also reported (e.g. Figure 2(d)). For each bin width, this value is derived by averaging the three observation time–accuracy curves for each SAR at that bin width, then choosing the maximum value in this average. In addition, high- vs. low-threshold SARs were denoted in the Results.

### 3. Results

Data were collected from eight SARs of which six were classified as high threshold and two as low threshold. The minimum ISIs among the high threshold SARs ranged from 13.3 to 22.5 ms, whereas those from the low threshold SARs were 7.5 and 12.6 ms, and these set the limit for maximum bin width for each cell. Figure 2 shows the relationship of maximum classification accuracy as a function of bin width using the JPBM and both alignment strategies for a typical high threshold SAR, where bin size is normalized by its minimum ISI for each neuron. Figure 2 illustrates the general finding (in all SARs) that maximal classification accuracy tends to increase with bin size up to the neuron’s corresponding minimum ISI, although minor fluctuations are evident. Maximal classification accuracy as a function of bin width using the EDBM and both alignment strategies for all SARs was also computed, and the same relationship was evident (data not shown). The JPBM performs better than chance (0.333 . . . ) at all bin widths and stimuli analysed. To compare the results derived from the JPBM with those derived from the
Figure 1. Illustration of JPBM and EDBM. (a) Model responses aligned by stimulus onset for 9, 12 and 15 ml stimuli. Waveforms show average TP inflation pressure and correspond to right-hand y-axes. Aligned underneath these histograms are a single test response (raw recording, ‘test response’) and its associated digital representation (‘test’). Beneath this are the probabilities of observing the same events in the three models as in the test response, and their final products. Note that the response is classified as belonging to the 15 ml model (check mark). (b) Graphical illustration of the EDBM, restricted to $n = 3$ dimensions for clarity. Each dimension provides the probability, $P$, of observing a spike in each bin. The locations of three hypothetical models ($M_1, M_2, M_3$) and a single hypothetical response (at $[0,1,0]$) are given. Euclidean distances ($d_1, d_2, d_3$) to each model from the test response are calculated, and in this case the response would be classified as belonging to model 1.
EDBM, the differences between the overall maximal classification accuracies produced by these two methods is given in Figure 3. Since most differences are positive, the classification accuracy using the JPBM is better than that of the EDBM in almost all SARs except for one high threshold and one low threshold SAR. There is no statistically-significant trend in the difference between different classification methods as a function of the bin width, implying that the general superiority of the JPBM method is not dependent on temporal resolution or alignment strategy. Over all SARs, the overall maximum accuracy for the JPBM method averaged 0.745 and 0.754 for the stimulus- and

Figure 2. Maximum classification accuracy for a typical high threshold SAR using JPBM. Classification accuracy is given as a function of normalized bin width for 9 ml (a), 12 ml (b), 15 ml (c) and overall (d) stimuli. Solid and broken traces are results from stimulus-and response-based alignments, respectively. Axes labels apply to all plots.

Figure 3. Comparison between JPBM and EDBM in overall maximal classification accuracy for all SARs using both alignment strategies. The differences in accuracy between the two methods (JPBM–EDBM) are shown for high threshold (solid traces) and low threshold (broken traces) SARs using stimulus-based (a) and response-based (b) alignment.
response-based alignments, respectively, and these were not statistically-significantly different ($P > 0.25$).

The effect of the number of bins (i.e. observation time) used in classifying spike trains on accuracy was also analysed. The results from one high threshold SAR using the JPBM and stimulus-based alignment are shown in Figure 4. Figure 4(a) (9 ml inflation) shows a relatively monotonic increase in accuracy as more bins (longer observation times) are used in the analysis, with accuracy increasing quickly up to 400 ms. Other features are evident in the observation time–accuracy relationships for 12 and 15 ml stimuli. Regardless of the details, one trend is consistent: the positive effect of increasing observation time on classification accuracy is strongest when bin width is just below the minimum ISI (13 ms for the SAR in Figure 4).

The accuracy data for all SARs using the JPBM and stimulus-based alignment revealed a similar trend. Maximum classification accuracies as a function of observation time, using bin widths just below the minimum ISI for all SARs, are plotted in Figure 5. The JPBM reveals that classification accuracy increases dramatically from the rising edge to TP peak, while passive deflation (after $\sim 600$ ms) contributes only modest improvements in accuracy compared to periods of active inflation. The EDBM, using the same bin width and alignment strategy, produces qualitatively identical results (data not shown).

4. Discussion

In this study, we employed the JPBM to classify individual SAR spike train responses using two different alignment strategies. We also compared it with the conventional EDBM.
Neither method assumes any response pattern (distributions) *a priori*. In general, the JPBM outperformed the EDBM on our SAR test dataset. Nevertheless, it is surprising that the maximum accuracy curves are similar in shape given the fact that their metrics are based on fundamentally different computations. Moreover, it is surprising that the two alignment strategies do not produce (statistically) significantly different results in these two methods. This implies that there is an internal consistency of the response sets to a given stimulus and that no systematic bias is present in the two analytical methods. The similarity of results produced by the two alignment methods is in accord with physiological requirements that the response reflects the stimulus features, since information about the stimulus amplitude and time course, rather than about a canonical model response, must ultimately be transmitted to the central nervous system by a given SAR spike train.

In general, increasing temporal resolution (i.e. decreasing bin widths) reduces classification accuracy regardless of which combination of method and alignment strategy is applied. Both methods show better performance when the bin size approaches the minimum ISI in the testing data. Theoretically, increasing temporal precision below that needed to classify response bins as having zero or one spike (i.e. a digital representation) may provide a more precise representation of response patterns, but this also increases the uncertainty associated with events in a particular bin. Thus, excessively high temporal resolution produces aberrant spike time jitter, resulting in higher classification error rates. This observation implies that dividing the silent epochs into many bins reduces classification capability while increasing computational burden of these methods.

Figure 5. JPBM-derived maximum accuracy as a function of observation time for all SARs. Accuracy values (left y-axes) are provided using stimulus-based alignment at bin widths just below the minimum ISI for six high threshold SARs (solid traces) and two low threshold SARs (dash traces) for test responses taken from 9 ml, 12 ml, 15 ml, and overall data sets ((a)–(d), respectively). Tracheal pressure waveforms (thick short dash traces, right y-axes) for representative stimuli in (a)–(c) are included in each plot to show the lung distention time course. Axes labels in (c) apply to all plots.
From the results presented, we conclude that the SAR response to the rising edge and peak of lung inflation contributes more information than passive deflation, since accuracy is only modestly improved during the latter epochs of the stimuli, if at all (Figures 4 and 5). This implies that information transmitted by SAR patterns is not uniform over different stages of the respiratory cycle. That is, active inflation produces responses that are more informative about lung distension amplitude, as the rate of rise in classification accuracy is the greatest then.

Finally, we note that worst classification performance was always worst for responses to 12 ml inflations (Figure 2(b)). This is because individual responses that deviated in either ‘higher’ or ‘lower’ from the 12 ml model in parameter space could be mistakenly classified as 15 or 9 ml, respectively, while the individual responses that were ‘above’ or ‘below’ the 15 or 9 ml models were still closest to those response models, and thus classified correctly. Thus, ceiling and floor effects were evident in this example of the classification problem, and indeed all applications where different stimuli are located in a continuum.

5. Conclusion
We developed and applied the JPBM to classify individual SAR responses into one of three different response categories evoked by one of three different stimuli. We tested its performance and compared it with the conventional EDBM. It performed better than the EDBM, with overall accuracy better than 70% for all SARs in our triad test. We explored the effects of varying temporal resolution, different alignment strategies, and length of responses on classification accuracy. We found that there is no statistically significant difference between the stimulus-based alignment and the response-based alignment when applying either the JPBM or EDBM to our dataset. We found that the increasing bin width from 1 ms to the minimal ISI increased classification accuracy. Finally, we also found that accuracy improves with increased observation time, but principally during active periods of lung inflation.

Acknowledgements
This work was supported in part by NIH R01 HL68143 (to RR).

References


