INSTANTANEOUS RATE ESTIMATION OF NEURONAL POINT PROCESSES FROM A COMPRESSED REPRESENTATION OF THEIR NONBINARY SPIKE TRAINS

Mehdi Aghagolzadeh [1] *Karim Oweiss* [1, 2]

[1] Electrical & Computer Engineering Department [2] Neuroscience Program
Michigan State University, East Lansing, Michigan

ABSTRACT

Estimating rate functions underlying neural point processes is essential for characterizing the firing patterns of cortical neurons involved in sensory and motor processing. This paper introduces a new method for directly estimating neuronal firing rates from a compressed representation of their extracellular recordings. The approach is based on extending a near-optimal sparse representation of the extracellular recordings to time scales matching those of the underlying rate functions, thereby performing the same role as a kernel-density estimator but at a much lower computational cost. Experimental results demonstrate that this method achieves comparable performance to the conventional Gaussian kernel estimation methods at a fraction of the computational cost. This makes it suitable to Brain Machine Interface applications where real time decoding of neural activity for neuro-motor prosthetic control using firing rate estimates is strongly desired.

***Index Terms*—** Brain machine interface, compressed sensing, spike train, spike sorting, firing rate.

1. INTRODUCTION

Decoding information in neuronal spike trains is a fundamental goal in systems neuroscience in order to better understand the complex mechanisms underlying brain function. In motor systems, these spike trains were demonstrated to carry important information about movement intention and execution of paralyzed subjects [1], and were shown to be useful in the development of neuro-motor prosthetic control of artificial limbs [2].

Cortically-controlled brain machine interface (BMI) systems rely fundamentally on instantaneous decoding of spike trains from motor cortical neurons. This decoding process is typically a cascade of processing steps illustrated in the top row of Fig.1. After preprocessing neural recordings through amplification, noise filtering and analog to digital conversion, spike detection and sorting followed by rate estimation is implemented to estimate the firing rate of individual neurons prior to decoding. The recorded data typically demands an ultra high bandwidth to permit the spike detection and sorting steps to take place with massive computational power [3]. Data compression is highly desired for these systems to be fully implantable and feature wireless communication with the outside world without compromising critical information needed for rate estimation downstream.

Estimating the firing rates from the set of event times post spike sorting is typically achieved by binning the data into time bins of equal width, and counting the number of events occurring within each bin. The resulting spike counts, often referred to as a rate histogram, constitute an instantaneous firing rate estimate. This is equivalent to convolving the spike train with a fixed-width rectangular window. This approach assumes that variations in the rate pattern over the bin width do not carry information. Alternatively, the firing rate can be also estimated by filtering the spike trains using a variable-width kernel function (e.g. a Gaussian) to yield smoothed estimations of the firing rates [4], where the temporal support of the kernel function is known to strongly impact the rate estimator [5].

In this paper, we show that a sparse representation of the neural signals, previously shown to yield substantial denoising and compression properties [6] enables adequate estimation of neuronal firing rates without the need to decompress, reconstruct and sort the spikes in the traditional sense. This is illustrated in the bottom half of Fig.1, where rate estimation and decoding of neural discharge patterns can be directly performed using the compressed data.

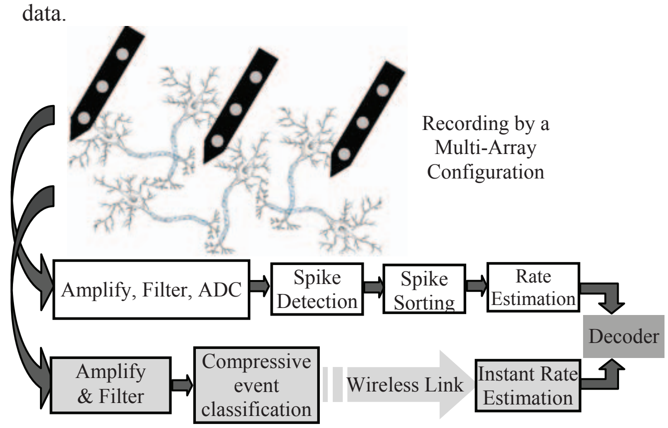


Fig.1. Schematic diagram of a typical data flow in a Brain Machine Interface application.

2. THEORY

In a typical neural recording experiment, the observations of interest are the times of events from a population of neurons. The set of event times that expresses the discharge pattern of an arbitrary neuron p can be modeled as a realization of an underlying point process with conditional intensity function -or firing rate- $\lambda_p(t|F)$. This intensity function is conditioned on some set, F, of intrinsic properties of the neuron itself and the neurons connected to it, and some extrinsic properties such as the neuron's tuning characteristics to external stimuli features [7]. Spike trains from motor cortex neurons can be modeled as a variant of the cosine tuning model of the neuron's preferred direction, θ_p, [8] as

$$\lambda_p\left(t_k \middle| \mathrm{x}_p\right) = \exp\left(\beta_p + \delta_p \cos\left(\tfrac{\theta(t_k)-\theta_p}{\omega_p}\right)\right) \quad (1)$$

where β_p denotes the background firing rate, $\theta(t_k)$ denotes the actual movement direction, and $x_p = [\theta_p, \delta_p, \omega_p]$ is a parameter vector governing the tuning characteristics of the modeled neuron. These are the tuning depth, δ_p, the preferred direction, θ_p, and the tuning width, ω_p. Event times can be modeled using (1) as the intensity function of an inhomogeneous Poisson process. The tuning term incorporates a neuron-dependent tuning width term, ω_p, an important parameter that affects the bin width choice for rate estimation. Variability in this term results in firing rates that are more stochastic in nature [9].

2.1. Sparse Representation of Spike Recordings

For compression purposes, it was shown in [5] that a carefully-chosen sparse transformation operator, such as a wavelet transform, can significantly reduce the number of coefficients representing each spike waveform recorded. The basis should be selected to keep the smallest number of coefficients without significant loss of spike features. To minimize the number of most important coefficients per spike event, ideally to a single feature, we note that the magnitude of the coefficients carry information related to the degree of correlation of the spike waveforms to the chosen basis [10]. Therefore, this information can be used to single out "the most significant" coefficient via a thresholding process. This *sensing* threshold is selected to preserve the ability to discriminate neuron p's events (alternative hypothesis) from those belonging to other neurons (null hypothesis) in a given node without necessarily minimizing a spike reconstruction error in a classical mean square error sense. This can be cast as a binary hypothesis testing problem in which

$$\left|g^j[k]\right| \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_p^j \qquad (2)$$

where γ_p^j is the sensing threshold of neuron p at time scale j and $g^j = \langle g_p, w_j\rangle$ is the convolution with the wavelet basis, w_j.

Fig.2 shows the wavelet decomposition tree for a set of five neurons recorded in vivo. Node 0 is the separation obtained in the time domain displayed in a 2-D feature space obtained from projecting spike events onto the largest eigenvectors. For Nodes 4 and 6, a sensing threshold can isolate units 4 and 1 respectively, while such a threshold is relatively harder to identify for Node 2. This is equivalent to a linear decision boundary of a 2-class problem. To detect the remaining classes, each detected class at a given node is first removed, then the decomposition proceeds to the next level and the 2-class binary decision process repeats until all nodes have been processed.

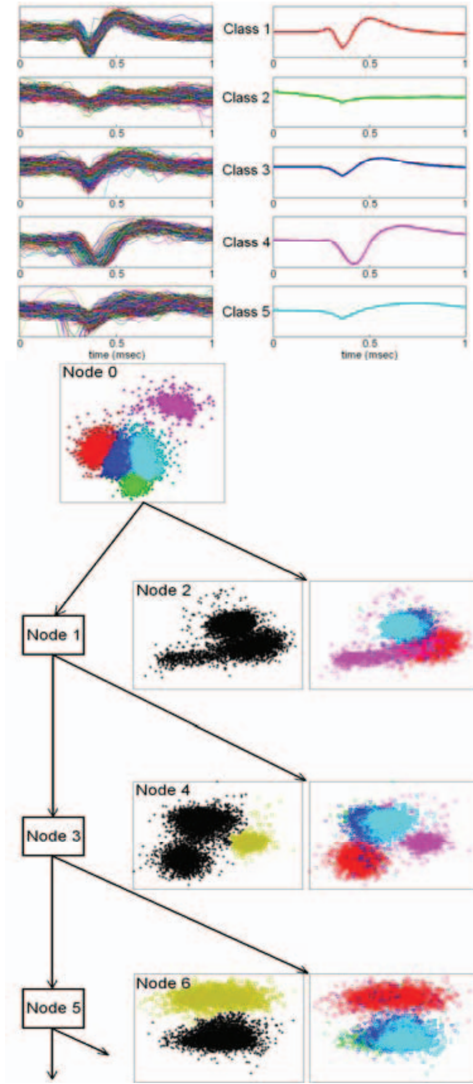


Fig.2. *Top:* Events from five recorded neurons aligned and superimposed on top of each other, and the corresponding spike templates obtained by averaging events from each neuron on the left column. *Bottom:* Wavelet decomposition tree showing the unit isolation quality, displayed in the 2-D feature space in every node. Verification of the algorithm is achieved by matching the gold cluster isolated by the sensing thresholds to a single colored cluster on the right side, which represent the remaining units.

2.2. Instantaneous Rate Estimation

The sparse representation obtained post thresholding provides a binary spike train of neuron p that can be expressed as

$$s_p = \sum_{i\in\{t_p\}} \sum_{k=0}^{N_s-1} g_p[k]\delta[i-k] \qquad (3)$$

where $\delta(.)$ is the Dirac delta function and g_p denotes a spike waveform generated by neuron p, where the length of each spike is N_s samples and $\{t_p\}$ is the set of event times. The compressed and thresholded representation can be expressed as

$$\bar{s}^j = \sum_{k^*\in\{t_p\}} g^j[k^*]\delta[i-k^*] \qquad (4)$$

A fundamental property of the sparse representation of the discrete wavelet transform suggests that as the level of decomposition increases, the wavelet coefficients become more representative of the intensity function rather than the temporal details of the spikes themselves. Mathematically, extending the wavelet decomposition to higher level is equivalent to convolving it with a wavelet kernel with increasing support. The temporal support, t_L, of the basis at level L is related to the sampling period, T_s, by

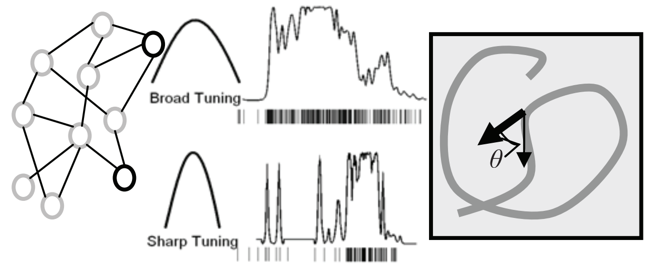
$$t_L = T_s n_w 2^{(L-2)} \qquad (5)$$

where n_w is the wavelet kernel support. For the symmlet4 basis used in this paper $n_w = 8$. Temporal characteristics of firing rate will be best characterized starting at level 6 and beyond where the basis support becomes long enough to include two or more consecutive spike events.

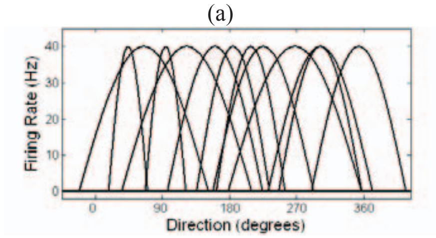
3. RESULTS

We used the rate estimation error in addition to the decoding performance as measures of success of the proposed method. A sample 2D arm trajectory, rate functions from two sample neurons with broad and sharp tuning widths, and their spike train realizations are shown in Fig.3a. Fig.3b illustrates the tuning characteristics of a subpopulation of 12 neurons used for encoding the 2D arm trajectory to demonstrate the heterogeneous characteristics of the model (1). A 9-second raster plot in Fig.3c illustrates the stochastic patterns obtained from the inhomogeneous Poisson model.

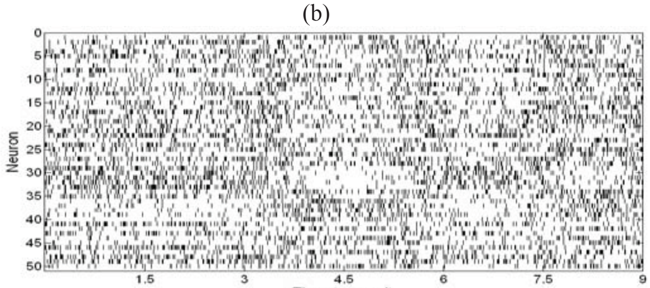
In Fig.4a, a 400 ms segment of the movement's angular direction over time is illustrated superimposed on the neuronal tuning range of five representative units with distinct tuning widths. The resulting firing rates and their estimators using the rate histogram (rectangular kernel), Gaussian kernel, and proposed DWT methods are illustrated for the five units, showing various degrees of estimation quality. As expected, the rate histogram estimate is noisy, while the Gaussian and Extended Discrete Wavelet Transform (EDWT) method perform better as indicated by a lower Mean Square Error (MSE)



(a)

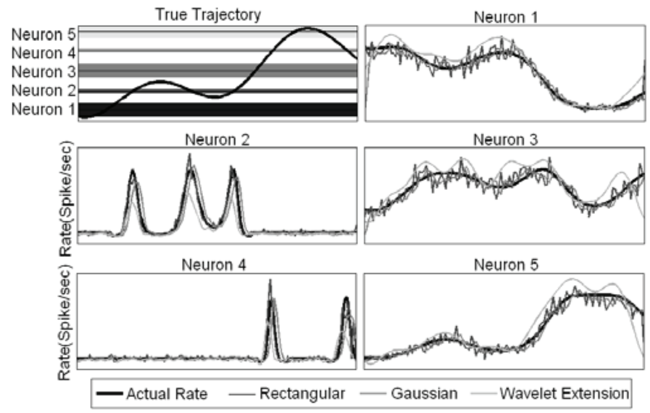


(b)

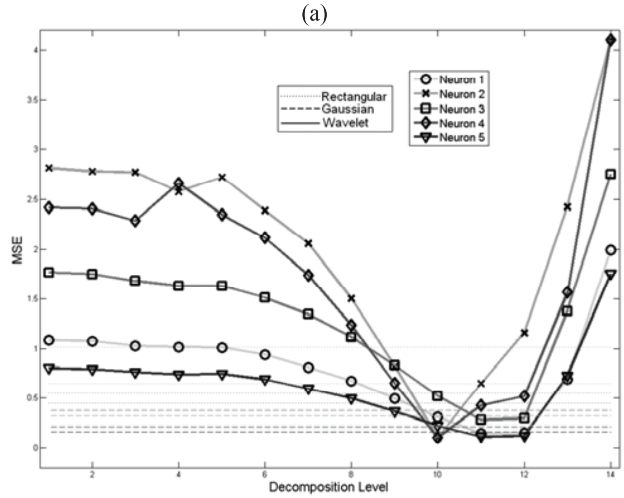


(c)

Fig.3. (a) Schematic of encoding 2D, non goal-directed arm movement. All neurons in the model were interconnected through random excitatory or inhibitory connections. (b) Tuning characteristics of a subset of the 12 neurons modeled with randomly chosen directions and widths. (c) Sample 9-second raster plot of spike trains from the population model.



(a)



(b)

Fig.4. (a) Top-left: 400 ms segment of angular direction of the movement trajectory superimposed on tuning characteristics of five representative units. The remaining figures represent the firing rate estimation for sharp and broad tuned neurons. (b) Mean square error between the actual and the estimated firing rate for each neuron with three firing rate estimation methods.

In Fig.4b, the relation between the wavelet kernel size and the MSE is quantified. As expected, decomposition levels with shorter kernel width (i.e., fine time scales) tend to provide the lowest MSE for neurons that are sharply tuned (10-levels). In contrast, a global minimum in the MSE is observed for broadly tuned neurons at coarser time scales, suggesting that these decomposition levels are better for capturing the time varying-characteristics of the firing rates (12-levels). Interestingly, the MSE for the EDWT method attains a lower level than both the rectangular and Gaussian kernel methods at the optimal time scale, clearly demonstrating the superiority of the proposed approach. Consequently, as the tuning broadens, larger kernel sizes (i.e. deeper decomposition levels) are required to attain a minimum MSE and thus better performance.

3.3. Decoding Performance

A sample trajectory and the decoded trajectory are shown in Fig.5 for four different cases: First, when no spike sorting is required. This is the ideal case in which every electrode records exactly the activity of one neuron, but is hard to encounter in practice. Second, when two or more units are recorded on a single electrode but no spike sorting is performed prior to rate estimation. Third, when spike sorting is performed for the later case using a traditional Principal Component Analysis (PCA) and Expectation Maximization (EM) clustering framework for spike sorting and the Gaussian kernel for rate estimation. And fourth, when combined spike sorting and rate estimation are performed using the proposed method. We used a linear filter for decoding in all cases. It is clear that the proposed method has a decoding error variance that is comparable to the third case, suggesting that the performance is as good as the standard method.

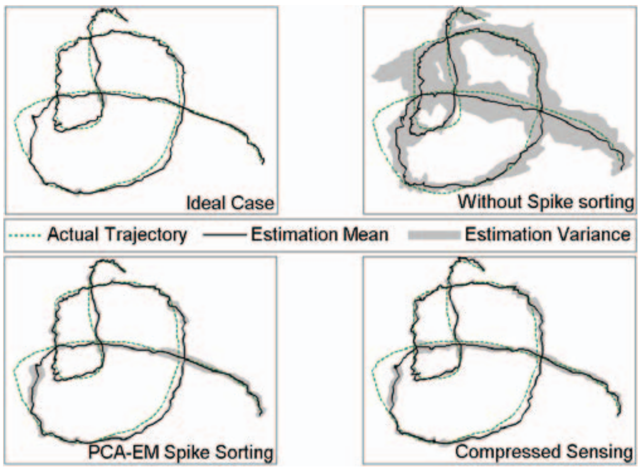


Fig.5. Decoding performance of a sample 2D movement trajectory. The black line is the average over 20 trials, while the gray shade represents the trajectory estimate variance. *Top left*: one neuron is recorded on any given electrode. The variance observed is due to the network term. *Top right*: every electrode records two neurons on average and no spike sorting is performed. *Bottom left*: standard spike sorting and rate estimation is implemented. *Bottom right*: Proposed method.

Comparing the cost of firing rate estimation through the standard time domain spike sorting/kernel smoothing approach and the compressed sensing approach shows substantial savings in computational costs for implantable neural prosthetic systems (Fig.6). This is mainly because the full reconstruction of the spike trains in the proposed approach is not necessary. On the other hand, the computational complexity of the PCA/EM/Gaussian kernel is attributed to the complexity in computing the eigenvectors of the spike data every time a new neuron is recorded, while in contrast, wavelets are universal, signal-independent approximators.

4. CONCLUSION

We have proposed a new approach to directly estimate instantaneous firing rates of cortical neurons from their compressed extracellular spike recordings. The approach is based on a sparse representation of the data and eliminates multiple blocks from the signal processing path in BMI systems. We showed that firing rates are estimated across a multitude of timescales, an essential feature to cope with the heterogeneous tuning characteristics of motor cortex neurons. We used the decoding of simulated 2D arm trajectories to demonstrate the quality of decoding obtained using this approach. The approach was compared to other methods classically used to estimate firing rates through a more complex processing path, and we demonstrated the improved performance attained with our approach.

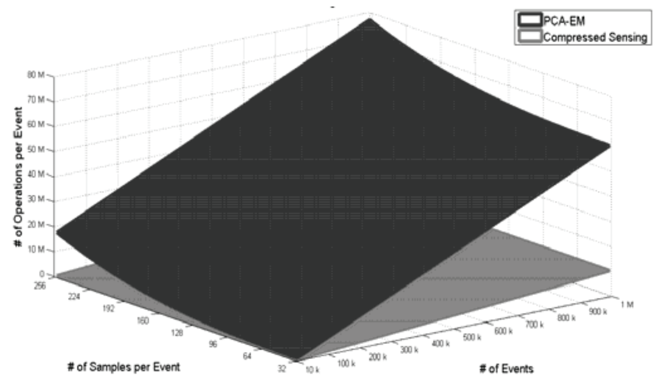


Fig.6. Computational complexity of the PCA/EM/Gaussian kernel and the proposed method: Computations per event displayed as a function of the number of events and number of samples per event.

5. REFERENCES

[1] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, “Neuronal ensemble control of prosthetic devices by a human with tetraplegia” *Nature*, Vol. 442, No. 7099., pp. 164-171, July 2006.

[2] M. Velliste, S. Perel, M. C. Spalding, A. S. Whitford, and A. B. Schwartz, “Cortical control of a prosthetic arm for self-feeding”, *Nature*, Vol. 453, pp. 1098-1101, June 2008.

[3] K. G. Oweiss, "A systems approach for data compression and latency reduction in cortically controlled brain machine interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 53, pp. 1364–1377, 2006.

[4] M. G. Paulin and L. F. Hoffman, "Optimal firing rate estimation," *Neural Network*, vol. 14, pp. 877-81, 2001.

[5] S. Cherif, K. E. Cullen, and H. L. Galiana, "An improved method for the estimation of firing rate dynamics using an optimal digital filter," *J Neurosci Methods*, 2008

[6] K. G. Oweiss and D. J. Anderson, "Tracking Signal Subspace Invariance for Blind Separation and Classification of Nonorthogonal Sources in Correlated Noise," *EURASIP Journal on Advances in Signal Processing*, 2007

[7] E. N. Brown, "Theory of Point Processes for Neural Systems " *in Methods and Models in Neurophysics*, C. C. Chow, B. Gutkin, D. Hansel, C. Meunier, and J. Dalibard, Eds. Paris: Elsevier, 2005, pp. 691-726.

[8] A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner, "Neuronal population coding of movement direction," *Science,* vol. 233, p. 1416, 1986.

[9] M. N. Shadlen and W. T. Newsome, "The Variable Discharge of Cortical Neurons: Implications for Connectivity, Computation, and Information Coding," *Journal of Neuroscience*, vol. 18, pp. 3870-3896, 1998.

[10] M. Aghagolzadeh, and K. Oweiss, “Compressed and Distributed Sensing of Neuronal Activity for Real Time Spike Train Decoding,” *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, 2009, to appear