Neuronal Spike Train Analysis in Likelihood Space

Yousef Salimpour,[1, 2] Hamid Soltanian-Zadeh,[1, 3, 4*]

Sina Salehi,[1, 2] Nazli Emadi,[1, 2] Mehdi Abouzari[1, 2]

[1]School of Cognitive Sciences (SCS),

Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

[2]Research Group for Brain and Cognitive Sciences,

School of Medicine, Shaheed Beheshti University, Tehran, Iran

[3]Center of Excellence for Control and Intelligent Processing,

Department of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

[4]Image Analysis Laboratory,

Department of Radiology, Henry Ford Health System, Detroit, Michigan, USA

[*]E-mail: hamids@rad.hfh.edu

Abstract

Background: Conventional methods for spike train analysis are based on rate function as a source of information while many experiments have shown a temporal coding mechanism too. Several techniques have been used for analyzing these two sources of information separately but using both sources in a single framework is still a challenging problem. Here, an innovative technique is proposed for spike train analysis by considering rate and temporal information.

Methodology/Principal Findings: Point process modeling approach is used to estimate stimulus conditional distribution based on observation of repeated trials. The extended Kalman filter is applied for estimation of the parameters in a parametric model. In order to extend this model from single neuron to neuronal population, marked point process strategy is used. Each spike train is transformed into a binary vector and then projected from the observation space onto the likelihood space. This projection generates a newly structured space that integrates temporal and rate information and thus improves performance of distribution-based classifiers. The stimulus-specific information is investigated in this space and used as a distance metric between two stimuli. To illustrate advantages of the proposed technique, spiking activity of inferior temporal cortex neurons of macaque monkey are analyzed in the observation and likelihood spaces. Based on goodness-of-fit, performance of the estimation method is demonstrated and the results are compared with the firing rate-based framework.

Conclusions/Significance: Due to improvement in neural discrimination of stimulus and new interpretation of information that can be used as a similarity measure, it is concluded that the likelihood space generates a more accurate representation of stimulus space. Finally, the problem of neural mechanism of visual object categorization may be addressed in this framework.

Introduction

Establishing quantitative correlation between neuronal spiking activity and external stimulus is a challenging task in neuroscience. Neurons generate series of spikes in response to the stimulus. A spike train is a stochastic process composed of a sequence of binary events that occurs in continuous time [1]. The point process theory is used as a stochastic framework to model the non-deterministic properties of the neural spike trains in which its parameters are estimated by recording the spike trains of a neuron in repeated trials [2]. The point process models can capture most of the nonlinear and stochastic properties of the neurons such as dynamic stimulus modulated response [3].

Point process framework is commonly used to model neuronal spiking activity [4]. This framework allows dynamic modeling, which is an important tool in computational neuroscience for studying neural dynamics [5]. Neural receptive field plasticity [6, 7], neural coding analyses [8, 9], neural spike train decoding [10, 11], neural prostheses [12, 13], analyses of learning [14, 15], analysis of neuronal spiking dynamic [16], and control algorithms' design for brain machine interfaces [17, 18], are examples of the neural dynamics. In most of the conventional methods, the neuronal firing rates of spiking activity are considered as a source of information and the temporal information is not included in the processing algorithms. In the case of temporal analysis in encoding the stimulus information, the neuronal rate functions are not considered. However, many experiments show different kinds of integration in temporal and rate information in encoding the stimulus features [19].

In many neuroscience research experiments, the aim of the study is to investigate how dynamic properties of neuronal system in the single or population level lead to the functional properties of specific brain regions [16]. The dynamic property of neural system, especially in spike train recording, indicates the requirement for dynamic signal processing methods. Despite the development of efficient dynamic signal processing algorithms, most current methods for neural spike train data processing are static and rate function based rather than dynamic and temporal based. For this reason, the enthusiasm for dynamic signal processing methods explicitly developed for neural spike train is increasing [20].

In this paper, a new feature space is generated by considering spike trains as binary vectors and projecting them onto the likelihood space. In this space, we are able to integrate temporal and rate information and compensate errors of modeling stimulus distribution in the observation space. These may improve performance of distribution-based classifiers by transforming the decision region into a contiguous region in the likelihood space.

Organization of the paper is as follows. First, we review point process modeling of neurons in terms of conditional intensity function and introduce the state space point process filtering approach by describing its parameter estimation method. Then, we show that the likelihood function of a spike train can be estimated based on the proposed model and also the likelihood space for each neuron can be generated by projecting its spike train. The marked point process is used for extending the model from a single neuron to a population of neurons. Properties of the likelihood space for spike trains are investigated. A new interpretation for information content of a spike train regarding a specific stimulus is introduced and used as a metric between the clusters of points in the projected space which is associated with the presented stimulus. Finally, we illustrate the efficiency of the estimation technique based on a goodness-of-fit criterion and demonstrate properties of the likelihood space by modeling neuronal spiking activity of the inferior temporal cortex of monkey in the single and population levels while performing a passive fixation task.

Materials and Methods

Point process modeling of a neuron: A stochastic neural point process can be completely characterized by its conditional intensity function which is a strictly positive function that gives a history-dependent generalization of the rate function of a Poisson process [21]. We use the conditional intensity function to characterize the spike train as a point process. We assume that in an interval $(0, T)$, J spikes are fired by a single neuron at times $t_1, t_2, \ldots, t_J$ for $t \in (0, T)$. The conditional intensity function is defined as:

$$\lambda(t|\theta(t), N_{0:t}) = \lim_{\Delta t \to 0} \left(\frac{Pr(N(t + \Delta t) - N(t) = 1|\theta(t), N_{0:t})}{\Delta t} \right) \tag{1}$$

where $\lambda(t|\theta(t), N_{0:t})$ is a conditional probability, $N_{0:t}$ includes the neuron's spiking history and the trace of spikes occurrences up to time t, and $\theta(t)$ is a parameter to be estimated. The $N(t)$ is the number of spikes fired by the neuron in $(0, t)$. Because the conditional intensity function completely defines the point process, to model the neural spike train in terms of a point process, it suffices to define its conditional intensity function. Parametric models may be used to express the conditional intensity as a function of covariates of interest [22].

In order to represent the point process model, we discretize the time interval $(0, T)$ by dividing it into K intervals of width $\Delta = TK^{-1}$ such that there is at most one spike per interval. For $k = 1, \ldots, K$, let ΔN_kbe the indicator of a spike in the interval $((k-1)\Delta, k\Delta)$, which is one if there is a spike and zero otherwise. We let $N_{1:k} = [\Delta N_1, \ldots, \Delta N_k]$ denote the spiking activity and denote the conditional intensity function $\lambda^S(k\Delta|\theta_k, N_{1:k-1})$ for the repeated trials when stimulus S is presented. The likelihood of a neural spike train is defined by finding the joint probability density of the data. It is shown that the joint probability of any point process is derived from the conditional intensity function by considering as a product of conditionally independent Bernoulli events [21]. If again we assume that on an interval $(0, T)$, J spikes are fired by a single neuron at times $t_1, t_2, \dots, t_J$ for the stimulus *S*, then the probability density of exactly these *J* spikes in $(0, T)$ is:

$$P(N_{0:T}\,|S) = exp\left[\sum_{k=0}^{K} log(\,\lambda^S(k\Delta|\theta_k, N_{1:k-1})\,\Delta)\Delta N_k\; - \sum_{k=0}^{K} \lambda^S(k\Delta|\theta_k, N_{1:k-1})\,\Delta\right]. \qquad (2)$$

We can evaluate the probability or the likelihood that the spike train comes from stimulus S by calculating the value of $P(N_{0:T}\,|S)$ using Equation (2). In this evaluation, the temporal pattern of spike train weighted by conditional intensity function is used [22]. In the rest of the paper, we use the marked point process to generalize the Equation (2) from single neuron to the population level.

Projection of spike trains onto the likelihood space: If there are *P* stimuli, any observed spike train $N_{0:T}$ must be related to one of the *P* stimuli S_1, S_2, ..., S_P. Let $P_N(N_{0:T}|S_1)$, $P_N(N_{0:T}|S_2)$, … , $P_N(N_{0:T}|S_P)$ represent the true distributions of the spike trains from the *P* stimuli. Let $\tilde{P}_N(N_{0:T}|S_1)$, $\tilde{P}_N(N_{0:T}|S_2)$, … , $\tilde{P}_N(N_{0:T}|S_P)$ be estimates of the true distributions. The likelihood projection of a sample path of spike train is defined as the operation $\mathrm{O}_P(N_{0:T})$ resulting in a *P*-dimensional likelihood vector, $Y_{N_{0:T}}$, as shown in Equation (3).

$$Y_{N_{0:T}} = O_P(N_{0:T}) = \;[log(\,\tilde{P}_N(N_{0:T}|S_1))\dots log(\,\tilde{P}_N(N_{0:T}|S_P))] \qquad (3)$$

The distributions $\tilde{P}_N(N_{0:T}|S_1)$, $\tilde{P}_N(N_{0:T}|S_2)$, … , $\tilde{P}_N(N_{0:T}|S_P)$ are the projecting distributions and the *P*-dimensional space whose coordinates are $log(\tilde{P}_N(N_{0:T}|S_1))$, $log(\tilde{P}_N(N_{0:T}|S_2))$, …, $log(\tilde{P}_N(N_{0:T}|S_P))$ is the likelihood space: When the dimension of the observation vector $N_{0:T}$ is greater than *P*, the likelihood projection operation $O_P(N_{0:T})$ is a dimensionality reducing operation (Figure 1) [23].

Properties of spike trains in the likelihood space: By construction of the likelihood space from spiking activity of the neurons, a categorization of stimulus is achieved by clustering of the projected neural data and decoding of the stimulus from the spike train. This can be considered as a distribution-based classification problem. Likelihood vector representations have the following properties that relate to clustering and classification in the likelihood space.

First, each spike train is assumed as a binary vector that contains temporal information in addition to the rate information, If for instance on an interval $(0, T)$, J spikes are fired by a single neuron at times $t_1, t_2, \dots, t_J$ for the stimulus S_i with conditional intensity function λ^{S_i} we can reinterpret the Equation 2 which is indicating the dependency of the components of the likelihood vector to temporal arrangement of the spikes that is weighted by the value of the conditional intensity.

Second, the projecting distributions represent a set of decision boundaries in the observation space that partition it into P decision regions. The decision region D_i for stimulus S_i is the region defined by:

$$N_{0:T} \in D_i \; if \; P(S_i)\tilde{P}_N(N_{0:T}|S_i) > P(S_j)\tilde{P}_N(N_{0:T}|S_j) \tag{4}$$

where $P(S_i)$ represents the a priori probability of stimulus S_i. The decision regions defined by Equation (4) may consist of several disconnected regions [23]. In the likelihood space, these regions are projected onto a region E_i defined by:

$$Y_{N_{0:T}} \in E_i \; if \; Y_{N_{0:T}}^{(i)} + \log(P(S_i)) > Y_{N_{0:T}}^{(j)} + \log\left(P(S_j)\right) \tag{5}$$

Equation (5) shows that if $Y_{N_{0:T}^{a}}$ and $Y_{N_{0:T}^{b}}$ both lie within E_i then, $\alpha Y_{N_{0:T}^{a}} + (1-\alpha) Y_{N_{0:T}^{b}}$ lies in E_i for any $0 \leq \alpha \leq 1$, thereby proving that the region E_i is convex and therefore connected.

Finally, in the observation space, the optimal minimum-error Bayesian classifier is given by the rule that $N_{0:T}$ is classified as belonging to the stimulus S_i such that i indexes the stimulus with the largest value for $P(S_i)\tilde{P}_N(N_{0:T}|S_i)$ [23]. A classifier that uses estimated distributions can be equivalently stated in terms of log-likelihoods as $logP(S_i) + log\tilde{P}_N(N_{0:T}|S_i)$. Classification between any two stimuli S_i and S_j is done as Equation (6). By considering $B_{ij} = lo\,g\,P(S_i) - lo\,g\,P(S_j)$ and A_{ij}^T a vector of 1 in the i^{th}component and -1 in the j^{th} component and 0 in the other components, Equation (6) can be redefined in the likelihood space as Equation (7) which is a simple linear discriminant with a slope of unity.

$$N_{0:T} \in S_i\, if\;\; log\left(\tilde{P}_N(N_{0:T}|S_i)\right) - \; log\left(\tilde{P}_N(N_{0:T}|S_j)\right) \;\geq\; lo\,g\,P(S_j) - lo\,g\,P(S_i) \tag{6}$$

$$N_{0:T} \in S_i\, if\;\; A_{ij}^T Y_{N_{0:T}} \geq \; B_{ji} \tag{7}$$

It is thus possible to define a classifier in the likelihood space that performs identically to a Bayesian classifier based on the projecting distributions in the observation space. It follows that the performance of the optimal classifier in the likelihood space cannot be worse than that in the observation space. It also follows that if the projecting distributions are the true distributions of the stimulus, then the optimal classification performance in the likelihood space is identical to the optimal classification performance in the observation space [23].

Extended Kalman filtering of a point process: The state space point process filtering approach is used for optimal estimation of parameters. In this approach, the counting process $N(t)$ is used by an observation equation as:

$$N(t) = \int_0^t \lambda(\tau|\theta(\tau), N_{0:\tau})d\tau + \eta(t) \tag{8}$$

where $\eta(t)$ is a zero mean error process that is the residual between a point process and its expectation. We construct a discrete time version of the observation Equation (8) for a fine partition of the observation interval, linearize its expected value as a function of the state process by using the linear terms of a Taylor expansion about the one-step prediction mean, and add Gaussian white noise errors as Equation (9).

$$\Delta N_k = \lambda_k + \Delta\left(\theta_k - \theta_{k|k-1}\right)\nabla_\theta \lambda(\theta_{k|k-1}) \;\; + \upsilon_k \tag{9}$$

where in the Equation (9) $\nabla_{\theta}\lambda(\theta_{k|k-1}) \; = \frac{\partial\lambda}{\partial\theta} \; |_{\theta=\theta_{k|k-1}}$ and the Gaussian error term υ_k should be selected so as to have similar statistical properties of the observation distribution. The variance of the discrete time approximation to the point process model is $\lambda_k\Delta$, which is unknown. Since Δ is sufficiently fine the $\upsilon_k \sim N(0, \lambda_k\Delta)$ might be a good choice. The state equation in Equation (10) is the Gaussian linear stochastic system where $\omega_k$is a zero-mean Gaussian noise with covariance matrix Q_k.

$$\theta_k = \theta_{k-1} + \omega_k \tag{10}$$

We model the conditional intensity function in terms of the state process as

$$\lambda^s \, (k\mathit{\Delta}|\theta_k, N_{1:k-1}) = exp\,(\theta_k) \tag{11}$$

In this kind of modeling, the history dependency in spiking activity within a trial is defined in terms of state process and the spiking activity between trials is independent. The exponential function is used as a parametric model for conditional intensity to ensure that the λ^s is strictly positive [24].

We apply the Kalman filtering method for parameter estimation [24]. It follows from the theory of point processes, that by taking the discrete approximation of the joint probability density of the spike train on the specific interval $((k-1)\Delta, k\Delta)$, the probability mass function of the observation equation for our state-space model is defined as:

$$P(\Delta N_k|\theta_k, N_{1:k-1}) = exp(\Delta N_k \, log(\lambda_k \mathit{\Delta}) - \lambda_k \mathit{\Delta}) \tag{12}$$

We define $N_k = [\Delta N_{1k}, \dots, \Delta N_{Jk}]$ as all the observation in the interval $((k-1)\Delta, k\Delta)$ across all J trials, $N_{1:K}^{J} = [N_1, \dots, N_K]$ and $\theta = [\theta_1, \dots, \theta_K]$ are the parameters of interest.

A standard approach for formulating state-space estimation algorithms uses the Bayes' rule and Chapman–Kolmogorov equations [25]. For the model defined in Equation (9) and (10), the *a* *priori* pdf is obtained from Equation (13) and the *a posteriori* pdf is obtained from Equation (14).

$$P(\theta_k|N_{1:k-1}) = \int P(\theta_k|\theta_{k-1})P(\theta_{k-1}|N_{1:k-1})d\theta_{k-1} \tag{13}$$

$$P(\theta_k|N_{1:k}) = \frac{P(\Delta N_k|\theta_k, N_{1:k-1})P(\theta_k\,|N_{1:k-1})}{\int P(\Delta N_k|\theta_k)P(\theta_k|N_{1:k-1})d\theta_k} \tag{14}$$

Equations (13) and (14) are a recursive system for computing the posterior density $P(\theta_k|N_{1:k})$. The first term in the numerator of Equation (14) is the probability mass function of the observation process in Equation (12), the second term is the one-step prediction density defined in Equation (13) and the denominator is a normalizing constant that ensures that the posterior probability density integrates to one. The challenge of this problem is to evaluate Equations (13) and (14) for the observation and system models in Equations (9) and (10).

Let $\theta_{k|k-1}$ and $W_{k|k-1}$define the mean vector and covariance matrix of the Gaussian approximation in Equation (13), and $\theta_{k|k}$ and $W_{k|k}$ be the mean vector and covariance matrix of the Gaussian approximation in Equation (14). The state transition model in Equation (10) is sufficient to compute the one step prediction probability densities [26, 27],

$$\theta_{k|k-1} = \theta_{k-1|k-1} \tag{15}$$

we write the posterior probability in the interval $((k-1)\Delta, k\Delta)$ by applying a Gaussian approximation.

$$\begin{aligned} P(\theta_k|N_{1:k}) &\approx exp(\Delta N_k\, log(\lambda_k \Delta) - \lambda_k \Delta) \;\times\; exp\left(-\frac{1}{2}\left(\theta_k - \theta_{k|k-1}\right)^T \left(W_{k|k-1}\right)^{-1}\left(\theta_k - \theta_{k|k-1}\right)\right) \\ &\approx exp\left(-\frac{1}{2}\left(\theta_k - \theta_{k|k}\right)^T \left(W_{k|k}\right)^{-1}\left(\theta_k - \theta_{k|k}\right)\right) \end{aligned} \tag{16}$$

The maximum a posterior estimate of the state is defined by the $\frac{\partial logP(\theta_k|N_{1:k})}{\partial\theta_k}\Big|_{\theta_k={\theta_k}^{MAP}} = 0$ and this relation should be approximately true for any value of ${\theta_k}^{MAP}$[26, 28]. We can therefore choose any specific point to evaluate this expression.

Evaluating at ${\theta_k}^{MAT} = \theta_{k|k-1}$ and rearranging the Equation gives [27, 28],

$$\theta_{k|k} = \theta_{k|k-1} + W_{k|k}\frac{\partial log\lambda_k}{\partial\theta_k}\Big|_{\theta_{k|k-1}}(\Delta N_{\mathrm{k}} - \lambda_k\Delta) \tag{17}$$

Since $W_{k|k-1} = Cov[\theta_k - \theta_{k|k-1}^{MAP}]$, we have

$$W_{k|k-1} = W_{k-1|k-1} + Q_k \tag{18}$$

and also we know that, $W_{k|k} = Cov[\theta_k - \theta_{k|k}^{MAP}]$. In the Kalman filtering framework, the updated *a posteriori* covariance is

$$(W_{k|k})^{-1} = (W_{k|k-1})^{-1} + \left(\frac{\partial log\lambda_k}{\partial\theta_k}\right)^T(\lambda_k\Delta)\left(\frac{\partial log\lambda_k}{\partial\theta_k}\right)\Bigg|_{\theta_{k|k-1}} \tag{19}$$

Thus far, the Kalman filter is completely derived for point process observation [27].

Marked point process modeling of a population: We consider a population of P neurons, responding simultaneously to a presentation of a stimulus. Their response are denoted by a vector $N_{0:T}^{1:P} = \{N_{0:T}^1, N_{0:T}^2, \dots, N_{0:T}^P\}$ where $N_{0:T}^i$ represents the stochastic response of the i^{th} neuron to a stimulus**.** The stimulus state is denoted by the scalar variable S**,** which is discrete in our case and selected with uniform probability from a stimulus set.

In order to find the probabilistic model for the populations of neurons, we apply the marked point process theory. Let (t_1, κ_1), ..., (t_k, κ_j) be the observation of P neuron over the interval [0,T]. The t_k is the spike instant in the pooled trains and the k_j is the label of the neuron which fire at time t_k [21, 29]. The log likelihood function $Log\ \mathrm{P}(N_{0:T}^{1:P}|\mathrm{S})$ of such a realization may be expressed in the form of the marked point process.

$$Log\ \mathrm{P}(N_{0:T}^{1:P}|\mathrm{S}) = \sum_{k=1}^{K} Log\ \lambda_S\ (t_k, k_i)\Delta N_k - \sum_{k=1}^{K}\sum_{i=1}^{J} \lambda_S\ (t_i, k_i) \qquad (20)$$

In this assumption, the marked point process is the combination of two independent processes. The ground process which is the result of pooling all the spikes in the interval [0,T], and the marked process which is the result of observing the label of the fired neuron at any spike instant. The conditional intensity function of population $\lambda_S\ (t_k, k_i)$ can be written as Equation (21) where $\lambda_{Sg}\ (t_k)$ is the intensity of the ground process, and $\lambda_{Sm}(k_i\ |t_k)$ is the intensity of a mark process at given time *t*.

$$\lambda_S\ (t_k, k_i) = \lambda_{Sg}\ (\ t_k) \times\ \lambda_{Sm}(k_i\ |t_k) \qquad (21)$$

The conditional intensity of the ground process is modeled with the sum of the intensities of the neurons in the ensemble Equation (22). The mark process that determines to which neuron the spike time should be attributed is randomly sampled for each spike time from a multinomial distribution with probability parameter as indicated in Equation (23)

$$\lambda_{Sg}\ (t_k) = \sum_{i=1}^{N} \lambda_S^i(t_k) \qquad (22)$$

$$\lambda_{Sm}(k_i\ |t_k) = \lambda_S^i(t_i)\left[\sum_{j=1}^{P} \lambda_S^j(t)\ \right]^{-1} \qquad (23)$$

By inserting the Equations (22) and (23) in (21) and substituting Equation (21) in (20), the Log-Likelihood function for marked point process model of P neurons in the population while the neurons observing the stimulus S can be written as Equation (24) [21, 29].

$$Log\ \mathrm{P}\left(\mathrm{N}_{0:\mathrm{T}}^{1:\mathrm{P}}|\mathrm{S}\right) = \sum_{k=1}^{K} \log(\sum_{i=1}^{P} \lambda_S^i\ (t_k))\Delta N_k + \sum_{i=1}^{K} \log(\frac{\lambda_S^i(k_i\ |t_k)}{\sum_{i=1}^{P} \lambda_S^i(k_i\ |t_k)})\Delta N_k - \sum_{i=1}^{K}\ \sum_{i=1}^{P} \lambda_S^i(t) \quad (24)$$

The Equation (24) is an extension of Equation (2) and can estimate the probability of observing response vector $N_{0:T}^{1:P}$ for the populations of neurons.

Information theoretic interpretation of spike trains in the likelihood space: Suppose a single neuron is responding to the stimulus set $S_1, S_2, \dots, S_P$ with the distributions of the spike trains $P_N(N_{0:T}|S_1)$, $P_N(N_{0:T}|S_2)$, … , $P_N(N_{0:T}|S_P)$. For any set of observations $N_{0:T,1}^{S_i}, N_{0:T,2}^{S_i}, \dots, N_{0:T,J}^{S_i}$ while the stimulus S_i was presented, if we consider the vector projected onto the likelihood space and scale each component with the probability distribution of the responses averaged across stimuli $P_N(N_{0:T}) = \frac{1}{P}\sum_{i=1}^{P} P_N(N_{0:T}|S_i)$, we can write the expectation of the vector with respect to $P_N(N_{0:T}|S_i)$ as:

$$E_{P_N(N_{0:T}|S_i)}\left[log(\frac{P_N(N_{0:T}|S_1)}{P_N(N_{0:T})}) \dots log(\frac{P_N(N_{0:T}|S_P)}{P_N(N_{0:T})})\right]$$

$$= \frac{1}{log_2^e}\left[\, I_{P_N(N_{0:T}|S_i)}(S_1) \dots I_{P_N(N_{0:T}|S_i)}(S_P)\right] \tag{25}$$

$$I_{P_N(N_{0:T}|S_i)}(S_i) = \sum_{j=1}^{J} P_N\left(N_{0:T,j}\middle|S_i\right) log_2 \frac{P_N\left(N_{0:T,j}\middle|S_i\right)}{P_N(N_{0:T,j})} \tag{26}$$

where $I(S_i)$ is the information specifically conveyed about stimulus S_i which is a direct quantification of variability in the responses elicited by that stimulus, compared to overall variability. By projecting all observations onto the likelihood space and scaling each component to the average response, we can define the distance between two stimuli S_i and S_j with respect to two different spiking activity patterns $N_{0:T}^{S_i}$, $N_{0:T}^{S_j}$ in the repeated trials as a difference between the information specifically conveyed about the two stimuli.

$$D\left(S_i, S_j\right) = \left[I_{P_N\left(N_{0:T}^{S_i}\middle|S_i\right)}(S_i) + I_{P_N\left(N_{0:T}^{S_j}\middle|S_i\right)}(S_i)\right] - \left[I_{P_N\left(N_{0:T}^{S_j}\middle|S_j\right)}\left(S_j\right) + I_{P_N\left(N_{0:T}^{S_i}\middle|S_j\right)}\left(S_j\right)\right] \tag{27}$$

Goodness-of-fit tests: We use the time-rescaling theorem to construct a goodness-of-fit test for a neural spike data model. Given a point process with conditional intensity function $\lambda\,(t|\theta(t), N_{0:t})$ and occurrence times $t_1, t_2, \dots, t_J$ where$t_j \in (0, T)$, if we define $z_j =$

$exp(-\int_{t_{j-1}}^{t_j} \lambda(\tau)d\tau)$, then these z_j are independent, exponential random variables with rate parameter one [10]. A common approach to measuring agreement between the model and the data is to construct a Kolmogorov-Smirnov (KS) plot. The KS plot is a plot of the empirical cumulative distribution function (CDF) of the rescaled times against an exponential CDF. If the conditional intensity model accurately describes the observed spiking data, then the empirical and model CDFs should roughly coincide, and the KS plot should follow a 45° line. If the conditional intensity model fails to account for some aspect of the spiking behavior, then that lack of fit will be reflected in the KS plot as a significant deviation from the 45° line. Confidence bounds for the degree of agreement between a model and the data may be constructed using the distribution of the Kolmogorov–Smirnov statistic [30].

Multidimensional scaling: Multidimensional scaling is a set of data analysis techniques that display the structure of distance-like data as a geometrical picture. Each object or event is represented by a point in a multidimensional space. The points are arranged in this space so that the distances between pairs of points reflect the similarities among the pairs of objects. That is, two similar objects are represented by two points that are close together, and two dissimilar objects are represented by two points that are far apart. A dissimilarity matrix must be real and symmetric with zeros along the diagonal and positive values elsewhere. In this paper, the classical multidimensional scaling is implemented by constructing a 2-dimensional space using the eigenvectors of the dissimilarity matrix corresponding to the two largest eigenvalues [31].

Recordings and stimuli: In a preparatory aseptic surgery, a block for head fixation and recording chamber were anchored to the dorsal surface of the skull. The position of the recording chamber was determined stereotaxically referring to the magnetic resonance images (MRIs) acquired before the surgery. Action potentials of single cells were recorded extracellularly with tungsten electrodes (FHC, ME) from the IT cortex while the monkey was performing a fixation task. The electrode was advanced with an oil-driven manipulator (Narishige) from the dorsal surface of the brain through a stainless steel guide tube inserted into the brain down to 10-15 mm above the recording sites. Recording positions were evenly distributed at anterior 15-20 mm over the ventral bank of the superior temporal sulcus and the ventral convexity up to the medial bank of the anterior middle temporal sulcus with 1-mm track intervals as illustrated in Figure 2. The recording was not biased by response properties. The action potentials from a single neuron were isolated by an offline sorting algorithm.

Responses of each cell were recorded with stimuli presented in a pseudorandom order. The stimulus set was repeated 49 ± 2 (median, 50) times for each recording site. The sequence of stimuli changed randomly in each repetition, and also for different recording sites, to avoid any consistent interaction between successively presented stimuli. The stimuli were gray scale photographs of natural and artificial objects isolated on a gray background. The size of the larger dimension (vertical or horizontal) of each stimulus was ~7° of visual angle.

The monkey had to maintain fixation within ±2° of a 0.5° fixation spot presented at the center of the display. The eye position was measured by an infra-red eye-tracking system (i_ rec, http://staff.aist.go.jp/k.matsuda/eye/), which allowed a precision of 1 deg or less for the measurement of eye position [32].

Ethics statement: All experimental procedures complied with the guidelines of the National Institutes of Health and the Iranian Society for Physiology. The use of non-human primates in this research was also in accordance with the recommendations of the Weatherall report, “the use of non-human primates in research”. All surgical procedures were performed under sodium pentobarbital anesthesia, and all efforts were made to ameliorate suffering of animals. The study protocol was approved by the ethics committee of School of Cognitive Sciences (SCS), Institute for Research in Fundamental Sciences (IPM) under permit number 08-06-83132001.

Results

In order to illustrate some of the properties of the likelihood space, the neural data of spiking activity from the inferior temporal cortex neurons of a macaque monkey is used. Each stimulus is presented for 300 ms and followed by 700 ms inter-stimulus blank interval. A 100 ms interval before stimulus presentation is recorded for the purpose of baseline activity study. Category selective neurons are entered in this study and the face selectivity is the most important feature for the neuron selection [32, 33]. Recording areas and the average firing rate's response of the neuronal population are illustrated in Figure 2.

Based on conditional intensity function model, point process filtering is applied and model parameters are optimally estimated. The stimulus effect in the conditional intensity format for the visual object is optimally estimated with 95% goodness-of-fit criteria as shown in Figure 3. The conditional intensity is used for the likelihood function estimation for each stimulus. The spike trains of the face selective neuron are projected onto the likelihood space (Figure 4). The dimension of the likelihood space is equal to the number of the stimuli; it can be created for any combinations of the stimulus set. Figure 4(A) shows the projection of the neural activity of the inferior temporal cortex when the human face and car pictures are presented to the animal. In Figure 4(B), a three-dimensional likelihood space is shown for the presentations of human face, dog face, and car to the same neuron.

In order to evaluate the "closeness" of the spike trains of the same stimulus after projecting them onto the likelihood space, the multidimensional scaling technique is applied to pair-wise comparisons of the entities. The multidimensional scaling allows us to visualize closeness of spike trains by representing them in a low-dimensional space [31]. The results of the multidimensional scaling analysis on normalized proximity matrices are shown in Figure 5. Figure 5(A) is an illustration of the same process for the spike train vectors before projection. Figure 5(B) shows the results of analysis on the distance measure in the likelihood space between any pairs of projected points in two dimensions. The Fisher's discriminant ratio is used to quantify the separability of the clusters in the two spaces. This criterion shows about 26% improvement in the separability of the clusters in the likelihood space on average. The amplitude of the difference between any pairs of vectors is defined as a distance in the observation space. The projected points represent each stimulus as a cluster. The clusters can be considered as estimates of representation of the neuron from stimulus space. The accuracy of the representation depends on the efficiency of the estimation method and the number of the spike trains observed.

We use the neural response of the 100 neurons recorded from the IT cortex of the monkey while doing the passive fixation task [33]. The spike trains of the neural ensemble in response to the human faces, cars, and dog faces for 50 repeated trials in 70-270 ms time intervals are modeled using marked point process, projected onto the likelihood space, and shown in Figure 6. By scaling each component with the response average and estimating the expectation of the component in the log-likelihood space, we estimate the stimulus specific information based on the center of each cluster. The center of the clusters are used for representing each stimulus category and the relative geometrical location of the cluster is considered as an interpretation of the neuronal population from the observed stimulus set.

In order to have a better comparison between the rate-based framework and temporal-based analysis, we use the neuronal activity of the same population in the same interval to extract similar result. We estimate the average firing rate of any individual neuron in 70-270 ms time interval then arrange them in a vector. The vectors are normalized and divided by their Euclidean lengths. We calculate a correlation-based distance measure and construct a relative geometrical interpretation of the different categories [34]. In Figure 7(A)-(B), the normalized distance measure based on similarity in the rate-based framework and the likelihood space framework are shown.

Based on the distance measurement and similarity, we use another analysis to compare the two frameworks. We use a 100 ms sliding time window with a step size of 10 ms and find the distance between two different categories in each step. In Figure 8, the distance or the dissimilarity between human face and car category is estimated in 100 ms sliding time window with 10 ms step size for the rate-based and the likelihood-space-based frameworks. We mark times of stimulus presentation and maximum distance occurrence in both frameworks. We use latency of maximum dissimilarity acutance as a criterion for temporal analysis of maximum information transmission.

Discussion

In this research, a new approach for analysis of spike trains is introduced where each spike train is considered as a binary vector and projected onto a lower-dimensional space. Many covariates are sources of spike generation in a single neuron and the observed spike trains are variable. The Kalman filtering based point process modeling approach and the state space generalized linear models help us to optimally estimate the conditional intensity function of the point process associated with each neuron for any stimulus. The time-rescaling theorem is used to construct goodness-of-fit tests for a neural spike data model. We model the spiking activity of the population of neurons using a single marked point process that has a conditional intensity which is the sum of the conditional intensities of all of the neurons in the population. To this end, class conditional distributions of stimuli are estimated and each observed vector is projected onto a specific point in the likelihood space.

The likelihood based approaches, which use the probability of neuronal response to a given stimulus, are widely used for fitting models and assessing their validity [35, 36]. They can be derived for several types of neural models and used for optimal decoding [37, 38]. In this study, we use the likelihood function to project spiking activity of neurons onto a new space, which might be a different application of the likelihood-based approach in spike train analysis. This is a new probabilistic interpretation of the spike train that enables us to apply the advanced signal processing and pattern recognition methods on neuronal data at the single neuron and population levels.

Projections of spike trains onto the likelihood space have important advantages. First, since each spike train or observed vector is directly used in the projection process, the temporal information ignored in the conventional methods is considered here. Secondly, the projected vectors are more separable in the likelihood space and also are less dependent on the accuracy of estimates of class conditional distributions thus, they may improve the performance of distribution-based classifiers. Finally, since the coordinates of the likelihood space are the stimuli conditional likelihood and the numbers of stimuli are less than the dimensions of the spike trains' binary vector, the projection is a dimension reduction process.

We introduce a new and novel interpretation of stimulus specific information conveyed by a neuron population. We extend the use of information theory to analyze spike trains by modeling the joint probability density function between the ensemble spiking activity and the biological signal explicitly and to compute stimulus specific information from the probability density function. In this approach, the direct parametric estimation of the conditional probability is used for the information estimation, which might be more accurate than rate-based approaches. We show that the expectation of each component in the likelihood space is proportional to the amount of information it conveys about a specific stimulus. Therefore, the difference between information content of the population about a specific stimulus can be considered as a distance metric and used for similarity measurement.

The limitations of the current study are: 1) the need for more observations compared to the conventional methods; 2) dependency of the model's accuracy in the population level to simultaneous observation of the neurons; and 3) complicate mathematics with more computational load with respect to conventional rate based spike train analysis such as peri-stimulus time histogram.

While this study establishes the feasibility of constructing likelihood space for the neuronal populations as a linear stochastic dynamical system with point process observation models, several extensions for the current framework are possible. First, there is a possibility to extend the current algorithm to a nonlinear state space model for computing smooth estimation of state estimate [39, 40]. Secondly, biophysically more realistic models can be used which are based on linear filtering stage followed by a noisy leaky integrate-and-fire spike generation mechanism [35, 36]. Thirdly, in our marked point process modeling of the population, we assumed that the neurons were independent given the value of the state process. Considering possible functional dependency among the neurons could broaden the application of the current framework to the various classes of the neuroscience problems. Fourthly, the emergence of multi-electrode arrays and the recent progress in multi-electrode recording enable us to interface with populations of neurons simultaneously [41, 42]. The marked point process modeling and likelihood space representation of the population might be applicable in real time observations such as neuro-prosthetic devices [43, 44]. Fifthly, a novel extension of the use of information theory to analyze multiple spike trains from developing probability models of joint spiking activity might be useful for investigating behavior of neuronal populations in dynamic stimulus coding. Finally, by collecting enough observation from the neuronal population, the representation of the population from the stimulus space may be demonstrated and the problems such as neural mechanism of stimulus categorization can be addressed in this framework.

Acknowledgements

The authors thank Dr. Hossein Esteky, the head of the School of Cognitive Sciences and the director of the Vision Lab, for providing experimental data and fruitful discussions.

Author Contributions

Conceived and designed the experiments: HS YS. Performed the experiments: YS SS NE MA. Analyzed the data: HS YS MA. Contributed reagents/materials/analysis tools: MA YS. Wrote the paper: HS YS MA.

References

1. Brown EN, Kass RE, Mitra PP (2004) Multiple neural spike train data analysis: state-of-the-art and future challenges. Nat Neurosci 7: 456-461.
2. Kass RE, Ventura V, Brown EN (2005) Statistical issues in the analysis of neuronal data. J Neurophysiol 94: 8-25.
3. Truccolo W, Eden UT, Fellow MR, Donoghue JP, Brown EN (2005) A point process framework for relating neural spiking activity for spiking history, neural ensemble and extrinsic covariate effects. J Neurophysiol 93: 1074-1089.
4. Brown EN, Nguyen DP, Frank LM, Wilson MA, Solo V (2001) An analysis of neural receptive field plasticity by point process adaptive filtering. Proc Natl Acad Sci U S A 98: 12261-12266.
5. Eden UT, Frank LM, Barbieri R, Solo V, Brown EN (2004) Dynamic analysis of neural encoding by point process adaptive filtering. Neural Comput 16: 971-998.
6. Frank LM, Eden UT, Solo V, Wilson MA, Brown EN (2002) Contrasting patterns of receptive field plasticity in the hippocampus and the entorhinal cortex: an adaptive filtering approach. J Neurosci 22: 3817-3830.
7. Frank LM, Stanley GB, Brown EN (2004) Hippocampal plasticity across multiple days of exposure to novel environments. J Neurosci 24: 7681-7689.
8. Barbieri R, Frank LM, Nquyen DP, Quirk MC, Solo V, et al. (2004) Dynamic analyses of information encoding by neural ensembles. Neural Comput 16: 227-308.
9. Brockwell AE, Rojas AL, Kass RE (2004) Recursive Bayesian decoding of motor cortical signals by particle filtering. J Neurophysiol 91: 1899-1907.

10. Ergun A, Barbieri R, Eden UT, Wilson MA, Brown EN (2007) Construction of point process adaptive filter algorithms for neural systems using sequential Monte Carlo methods. IEEE Trans Biomed Eng 54: 419-428.
11. Deneve S, Duhamel JR, Pouget A (2007) Optimal sensorimotor integration in recurrent cortical networks: a neural implementation of Kalman filters. J Neurosci 27: 5744-5756.
12. Yu BM, Kemere C, Santhanam G, Afshar A, Ryu SI, et al. (2007) Mixture of trajectory models for neural decoding of goal directed movements. J Neurophysiol 97: 3763-3780.
13. Shoham S, Paninski LM, Fellows MR, Hatsopoulos NG, Donoghue JP, et al. (2005) Statistical encoding model for a primary motor cortical brain-machine interface. IEEE Trans Biomed Eng 52: 1312-1322.
14. Smith AC, Frank LM, Wirth S, Yanike M, Hu D, et al. (2004) Dynamic analysis of learning in behavioral experiments. J Neurosci 24: 447-461.
15. Smith AC, Wirth S, Wendy AS, Brown EN (2007) Bayesian analysis of interleaved learning and response bias in behavioral experiments. J Neurophysiol 97: 2516-2524.
16. Czanner G, Eden UT, Wirth S, Yanike M, Suzuki WA, et al. (2008) Analysis of between-trial and within-trial neural spiking dynamics. J Neurophysiol 99: 2672-2693.
17. Srinivasan L, Brown EN (2007) A state-space framework for movement control to dynamic goals trough brain-driven interfaces. IEEE Trans Biomed Eng 54: 526-535.
18. Serruya MD, Hatsopoulos NG, Paninski L, Fellows MR, Donoghue JP (2002) Brain-machine interface: instant neural control of a movement signal. Nature 416: 141-142.
19. Kiani R, Esteky H, Tanaka K (2005) Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. J Neurophysiol 94: 1587-1596.
20. Dayan P, Abbott LF (2001) Theoretical Neuroscience. Cambridge, Massachusetts: MIT Press.

21. Daley D, Vere-Jones D (2003) An Introduction to the Theory of Point Process, 2nd ed, New York: Springer-Verlag.
22. Brown EN, Barbieri R, Eden UT, Frank LM (2003) Likelihood methods for neural data analysis. In: Feng J, editor. Computational Neuroscience: A Comprehensive Approach. London: CRC. pp. 253-286.
23. Singh R, Raj B (2004) Classification in likelihood spaces. Technometrics 46: 318-329.
24. Salimpour Y, Soltanian-Zadeh H (2009) Particle filtering of point process observation. In: Proc. 4th Int. IEEE/EMBS Conf. pp. 718-721.
25. Barbieri R, Frank LM, Nguyen DP, Quirk MC, Solo V, et al. (2004) Dynamic analyses of information encoding by neural ensembles. Neural Comput 16: 277-308.
26. Simon D (2006) Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches. Wiley-Interscience.
27. Haykin S (2002) Adaptive Filter Theory. Prentice Hall.
28. Grewal M, Andrews A (2008) Kalman filtering theory and practice using Matlab, 3rd ed, Wiley-IEEE.
29. Jacobsen M (2005) Point Process Theory and Applications: Marked Point and Piecewise Deterministic Processes. Birkhäuser Boston.
30. Brown EN, Barbieri R, Ventura V, Kass RE, Frank LM (2002) The time-rescaling theorem and its application to neural spike train data analysis. Neural Comput 14: 325-346.
31. Cox TF, Cox MAA (2001) Multidimensional Scaling. Chapman and Hall.
32. Kiani R, Esteky H, Mirpour K, Tanaka K (2007) Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. J Neurophysiol 97: 4296-4309.

33. Hung CP, Kreiman G, Poggio T, DiCarlo JJ (2005) Fast readout of object identity from macaque inferior temporal cortex. Science 310: 863-866.
34. Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, et al. (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293: 2425-2430.
35. Paninski L (2004) Maximum likelihood estimation of cascade point-process neural encoding models. Network 15: 243-262.
36. Paninski L, Pillow J, Simoncelli E (2004) Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. Neural Comput 16: 2533-2561.
37. Keat J, Reinagel P, Reid R, Meister M (2001) Predicting every spike: a model for the responses of visual neurons. Neuron 30: 803-817.
38. Pillow J, Paninski L, Uzzell V, Simoncelli E, Chichilnisky E (2005) Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. J Neurosci 25: 11003-11013.
39. Smith AC, Brown EN (2003) Estimating a state-space model from point process observations. Neural Comput 15: 965-991.
40. Godsill SJ, Doucet A, West M (2004) Monte Carlo smoothing for nonlinear time series. J Am Stat Assoc 99: 156-168.
41. Serruya MD, Hatsopoulos NG, Paninski L, Fellows MR, Donoghue JP (2002) Instant neural control of a movement signal. Nature 416: 141-142.
42. Wessberg J, Stambaugh CR, Kralik JD, Beck PD, Laubach M, et al. (2000) Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. Nature 408: 361-365.
43. Taylor DM, Tillery SI, Schwartz AB (2002) Direct cortical control of 3D neuroprosthetic devices. Science 296: 1829-1832.

44. Musallam S, Corneil BD, Greger B, Scherberger H, Andersen RA (2004) Cognitive control signals for neural prosthetics. Science 305: 258-262.

Figure Legends

Figure 1. Projection of spike train onto likelihood space. Sample response of the neuron to the stimulus presentation in the raster plot format. The transformation of the spike train from the observation space into a likelihood space is illustrated. Each point in the new space is the projection of the binary vector of spike train with respect to the stimuli conditional distribution.

Figure 2. Recording from inferior temporal cortex. Recording positions at anterior 15-20 mm over the ventral bank of the superior temporal sulcus and the ventral convexity up to the medial bank of the anterior middle temporal sulcus with 1-mm track intervals and the average firing rate's response of the neuronal population.

Figure 3. Model parameter estimation. Sample responses of a neuron from inferior temporal cortex of a macaque monkey while the animal is doing the passive fixation task. The raster plot and the conditional intensity function are shown for a human face and a car.

Figure 4. Projection onto likelihood space. (A) Reconstruction of likelihood space for the neural activity of the inferior temporal cortex, while the human face and car pictures are presented. (B) The likelihood space for the same neuron while presenting human face, dog face, and car pictures.

Figure 5. Multidimensional scaling in observation space and likelihood space. Multidimensional scaling technique is used to illustrate capability of the likelihood space in increasing the sreparability of the clusters. (A) The distance measurement and multidimensional

scaling results for the pairs of spike trains from the human face and car stimuli in the observation space. (B) The distance measurement and multidimensional scaling results for the same spike trains after projection them onto the likelihood space.

Figure 6. Population based likelihood reconstruction. The likelihood space representation for the populations of neurons while the human face, dogface, and car pictures were presented.

Figure7. Between stimulus distance measure. The likelihood space and correlation based representations of stimulus space for the populations of neurons while presenting of human face, dog face and car pictures. The normalized neural representation of distance in the correlation based (A) and likelihood space (B).

Figure 8. Dynamic between-stimulus distance measure. (A) Dynamic distance measurement between pairs of stimuli from two different categories in 100 ms sliding time window with 10 ms sliding step based on correlation distance. (B) Dynamic distance measurement for the same stimulus pair with 100 ms sliding time window and 10 ms sliding step based on stimulus distance in the likelihood space.