Information Theoretical Analysis of Instantaneous Motor Cortical Neuron Encoding for Brain-Machine Interfaces

Yiwen Wang1, Justin C. Sanchez2, Jose C. Principe1
1Department of Electrical & Computer Engineering, University of Florida, Gainesville, FL 32611, USA
2Department of Pediatrics, Neuroscience, and Biomedical Engineering, University of Florida, Gainesville, FL 32610, USA

E-mail: wangyw@cnel.ufl.edu

Abstract. Sequential estimation algorithms based on spike trains for motor Brain-Machine Interfaces (BMI) require knowledge of both neuronal representation encoding of movement and movement decoding from spike train activity. In these BMIs, an instantaneous encoding estimation is necessary which is unlike the methods commonly used that are based on time windows of neural and kinematic data. An online, instantaneous encoding analysis based on information theoretic techniques is developed using the relationship between the instantaneous kinematic vector and neural firing in the motor cortex. Moreover, mutual information is utilized as a tuning criterion to provide a way to estimate the optimum time delay between motor cortical activity and the observed kinematics. More than half (58.38%) of the neurons instantaneous tuning curves display a 0.9 correlation coefficient with those estimated with the temporal kinematic vector. Unlike the windowed methods, one of the characteristics of the instantaneous model is that it works within the dynamic range of the kinematics. This paper shows that the instantaneous estimation provides better encoding when compared with the window models in real experimental data.

1. Introduction
Brain-Machine Interfaces (BMI) are a framework in which the spatial and temporal structure of neural activity is used to directly control a prosthetic device. In this framework [Wessberg et al. 2000, Serruya et al. 2002], neuronal activity (local field potentials and single unit activity) is synchronously collected from microelectrode arrays implanted into multiple cortical areas while the animal is performing 3-D food-reaching tasks, or 2-D target-tracking tasks. Several signal-processing approaches have been applied to extract the functional relationship between the neural recordings and the animal’s kinematic trajectories [Wessberg et al. 2000, Sanchez et al. 2002, Kim et al. 2003, Wu et al. 2006, Brockwell et al. 2004]. The models predict movements and control a prosthetic robot arm or computer to implement them. Several approaches have been proposed to advance “black-box” modelling in BMIs [Sanchez & Principe 2007] by incorporating important properties of the motor nervous system into the model and then using data-driven methodologies to determine unknown parameters. One of the best well-known methods is the population vector algorithm proposed by Georgopoulos et al. [1986] and used in BMIs by Schwartz et al. [1988]. In this method, the movement direction is predicted from the preferred direction vectors of all neurons, which are appropriately weighted according to each neuron’s tuning curve. This basic idea was further improved using a state
model within a Bayesian formulation [Wu et al. 2006, Brockwell et al. 2004, Wang et al. 2007a]. The probabilistic approach based on Bayesian estimation constructs the posterior density of each kinematic state given the spike trains at each time step from the prior density of the state. The prior density in turn is the posterior density of previous time step updated with the discrepancy between an observation model and the spike train. The observation model linking the measurement of the noisy neural activity to the kinematics implicitly utilizes the tuning characteristics of each neuron. In our newly proposed Monte Carlo sequential estimation algorithm operating directly on point processes [Wang et al. 2006], the Bayesian approach analyzes and infers the kinematics as a state variable of the neural dynamical system without the constraints of linearity and Gaussianity. Accurate modeling of the neuron tuning properties in the observation model is critical to decode the kinematics by expectation of the posterior density or by maximum likelihood estimation.

The tuning, also called the encoding function, mathematically models how a neuron represents behavioral responses or how it responds to a stimulus. The parameterization of a tuning function requires an understanding of three interconnected aspects: 1) What is the behavior/stimulus? 2) How does the neuron encode it? 3) What is the criterion for quantifying the quality of the response? For neurons located in the motor cortex, researchers first developed the static descriptions of movement-related activity by applying electrical stimuli to motor areas to elicit muscle contraction [Fritsch & Hitzig 1870, Leyton & Sherrington 1917, and Schafer 1900]. Later, movement direction was correlated with cortical firing in a center-out task where the tuning function was initially modeled as a cosine curve [Georgopoulos et al. 1982]. The peak discharge rate of a cell is called preferred direction. Researchers have also proposed a variety of parametric models to describe the motor representation in neural firing. Linear relationships from motor cortical discharge rate to speed and direction have been constructed [Moran & Schwartz 1999]. The linear filter takes into account the sensitivity of preferred direction, the position and speed of the movement to represent the firing rate in cortical activity [Roitman et al. 2005]. However, linear encoding captures only a fraction of the overall information transmitted because the neuron exhibits nonlinear behavior with respect to the input signal. Brown et al. [2001] used a Gaussian tuning function for hippocampal pyramidal neurons. Brockwell et al. [2003] assumed an exponential tuning function for their motor cortical data. These nonlinear mathematical models are not optimal for dealing with real data because the tuned cells could have very different tuning properties. Simoncelli and Paninski et al. [2004] further improved the linear idea and proposed a Liner-Nonlinear-Poisson (LNP) model to cascade the linear stage with a nonlinear transformation as the second stage, which gives a conditional instantaneous firing rate to the Poisson spike generating model at the third stage.

In the LNP model, the position or velocity at all relevant times within a temporal window was utilized to extract the information between neuronal activity and animal movement trajectories. During a continuous target tracking task, Paninski et al. [2004b] studied the temporal dynamics of M1 neurons for position and velocity of hand motion given the firing rate. The linear filter in the LNP model averages the temporal position or velocity within the window and so it smoothes the statistical curves on the stimulus distribution and provides the widely known exponential increasing nonlinearity that relates neuronal firing rate to the projected kinematics. Unfortunately, the averaging builds up an $N$ to $1$ temporal mapping between the kinematic variables (position or velocities) and the neural spikes that negatively impacts our goal of building sequential estimation algorithms. Indeed, sequential models require inferring the estimation of the kinematics from current neuronal spike times. Therefore, the instantaneous one to one functional tuning relationship between the kinematics and neuron activities are needed to decode the kinematics online and to avoid the error accumulation within the windowed kinematic vector. Moreover, the analysis of the receptive fields of motor cortex neurons is different from the stimulus-response analysis in sensory cortices, because there is always a time delay between neuron spiking and the movement response. This delay must be taken into consideration in BMI decoding algorithms. The estimation of instantaneous tuning parameters is more difficult and more prone to errors. Therefore, we will have to evaluate how effective the instantaneous estimator really is.
In the literature, mutual information has been used to differentiate the raw stimulus ensemble from the spike-triggered stimulus distribution [Simoncelli et al. 2004, Sharpee et al. 2002], as well as to estimate the minimal number of delay samples in the temporal kinematics needed to represent the information extracted by the full preferred trajectory of a given cell [Paninski et al. 2004b]. In this paper, we also apply an information theoretical analysis but in section 2 we modify the method to estimate the immediate tuning properties of motor cortical neurons as required by sequential estimation algorithms for BMIs based on spike timing. The information theoretical analysis is also applied to the projective nonlinear-Poisson encoding analysis to estimate the causal time delay. The nonlinearity of the instantaneous tuning curves is compared to the windowed kinematics method in section 3, followed by the conclusion and future work.

2. Data collection and methods

2.1. Data collection

The Brain-Machine Interface paradigm was designed and implemented in Dr. Miguel Nicolelis laboratory at Duke University. Chronic, neural ensemble recordings were collected from the brain of an adult female Rhesus monkey, and synchronized with task behaviours. Microelectrode arrays were chronically implanted in five cortices: right dorsolateral premotor area (PMD), right primary motor cortex (MI), right primary somatosensory cortex (S1), right supplementary motor area (SMA), and the left primary motor cortex (MI).

After the surgical procedure, a multi-channel acquisition processor (MAP, Plexon, Dallas, TX) cluster was used in the experiments to record the neuronal action potentials simultaneously. Analog waveforms of the action potential were amplified and band pass filtered from 500 Hz to 5 kHz. The spikes of single neurons from each microwire were discriminated based on time-amplitude discriminators and a principal component analysis (PCA) algorithm [Nicolelis et al. 1997, Wessberg et al. 2000]. The firing times of each spike were stored. Table 1 shows the assignment of the sorted neural activity to the electrodes for different motor cortical areas [Kim 2005].

<table>
<thead>
<tr>
<th>Monkey 1</th>
<th>Right PMd</th>
<th>Right MI</th>
<th>Right S1</th>
<th>Right SMA</th>
<th>Left MI</th>
</tr>
</thead>
<tbody>
<tr>
<td>(left handed)</td>
<td>1-66(66)</td>
<td>67-123(57)</td>
<td>124-161(38)</td>
<td>162-180(19)</td>
<td>181-185(5)</td>
</tr>
</tbody>
</table>

The monkey performed a two-dimensional target-reaching task to move the cursor on a computer screen by controlling a hand-held joystick to reach the target. The monkey was rewarded when the cursor intersected the target. The corresponding position of the joystick was recorded continuously for an initial 30-min period at a 50 Hz sampling rate, referred to as the “pole control” period [Carmena et al. 2003].

2.2. Information theoretical neural encoding

This section implements an information theoretical methodology to address instantaneous neuronal encoding properties. The analysis is based on a statistical procedure for quantifying how neuronal spike trains directly encode arm kinematics. All of the evaluation is performed directly with the neural spike times, which preserves the fine time structure of the representation without determining a rate code and its associated window size commonly chosen by the experimenter.

2.2.1 Instantaneous tuning function in motor cortex

The literature contains many different types of tuning functions, i.e. linear, exponential, Gaussian [Moran & Schwartz, 1999, Eden & Frank et al., 2004]. These nonlinear mathematical models are not optimal for dealing with the real data because each neuron very likely has different tuning properties.
The accuracy of the tuning function estimation will directly affect the Bayesian decoding approach and, therefore, the results of the kinematic estimation in BMIs. The spike-triggered average (STA) is one of the most commonly used white noise analysis methods [deBoer & Kuypers 1968, Marmarelis & Naka 1972, Chichilnisky, 2001], applicable when data is uncorrelated. It is applied for instance in the study of auditory neurons [Eggermont et al. 1983], retinal ganglion cells [Sakai & Naka 1987, Meister et al. 1994], lateral geniculate neurons [Reid & Alonso, 1995], simple cells in primary visual cortex (V1) [Jones & Palmer, 1987; McLean & Palmer, 1989, DeAngelis et al. 1993]. STA provides an estimate of the first linear term in a polynomial series expansion of the system response function with the assumptions that the raw stimulus distribution is spherically symmetric or elliptically symmetric (whitening operation is then necessary), and the raw stimuli and the spike-triggered stimuli distribute differently in terms of the mean. If the system is truly linear, STA provides a complete characterization. This linear approximation was improved by Simoncelli, Paninski and colleagues [Simoncelli et al. 2004]. By parametric model identification, the nonlinear property between the neural spikes and the stimuli was directly estimated from data, which is more reliable than just assuming a linear or Gaussian dependence. In our sequential estimation for BMI studies [Wang et al. 2006] it provides a very practical way to acquire the prior knowledge (the tuning function) for decoding purposes. This technique estimates the tuning function by a Linear-Nonlinear-Poisson (LNP) model [Simoncelli et al. 2004], which is composed of a linear filter followed by a static nonlinearity then followed by a Poisson model, as shown in Figure 1.

![Fig. 1 Block diagram of projected-nonlinear-Poisson model](image)

The linear filter projects the multi-dimensional kinematic vector into its weight vector \( \vec{k} \) (representing a direction in space), which produces a scalar value that is converted by a nonlinear function \( f \) and applied to the Poisson spike-generating model as the instantaneous conditional firing probability \( p(\text{spike} | \vec{k} \cdot \vec{x}) \) for that particular direction in the high dimensional space. In our work the optimal linear filter actually projects the multi-dimensional kinematic vector \( \vec{x} \) built from the position, velocity and acceleration in x and y along the direction where they differ the most from the spike triggered kinematic vectors. This projection could represent the transformation between kinematics to muscle activation [Todorov 2000]. The nonlinear function \( f \) represents the neuron nonlinear response, which accounts for all of the processing of the spinal cord and deep brain structures to condition the signal for activation operations [Todorov 2000]. The Poisson model, which encodes the randomness of neural behavior, generates spike trains with an instantaneous firing probability defined by the nonlinear output. This modeling method assumes that the generation of spikes depends only on the recent stimulus and is historically independent of previous spike times. Previous work [Paninski et al. 2004a, Paninski et al. 2004b] utilized a window in time approach to build a smoother statistical tuning function from temporal kinematics to instantaneous neural firing rate. In the encoding stage, the kinematic variable within a window that embeds temporal information before and after the current neuron firing time is used as a high dimensional input vector. The linear-nonlinear stage of the LNP model generates a one-dimensional output as the estimated firing rate for the Poisson stage. However, the sequential estimation model of our BMI requires just the opposite, i.e. we need to predict from the current neural activity event a sequence of kinematics, especially for the neurons in M1. When we infer the kinematics during a certain window length with respect to a particular spike, the state estimation error can accumulate easily as the estimation is recursively
propagated into the next time iteration to build the vector during the window. Thus, a one-to-one mapping between the instantaneous kinematics and the neural activities is of paramount importance for the online decoding purpose. The other issue is to estimate appropriately the optimal delay in the instantaneous functional mapping. Due to the decrease in the amounts of data, the instantaneous decoding is expected to be noisier (fewer data to identify the transfer function), but there are also possible advantages. Compared to the windowed method of Paninski et al. [2004b], instantaneous estimation works directly in the dynamic range of the kinematic signals instead of being affected by all the temporal information embedded within the window. To deal with the sensitivity issue for neural tuning identification, the method works with the full kinematic vector containing the instantaneous position, velocity and acceleration to include the information that each kinematic variable conveys for tuning, which ultimately is what is needed in BMI decoding.

Estimation of the instantaneous encoding depends upon the ability to estimate the appropriate time delay between motor cortical neuron activity and kinematics [Wu et al. 2006]. Due to the propagation effects of signals in the motor and peripheral nervous system and to preserve causality, the intended movement is executed after the motor cortical neuron fires. In the temporal kinematic encoding by LNP models, a window that usually samples 300 msec before and 500 msec after the current neural firing rate [Paninski et al. 2004b] is used to construct the high dimensional kinematic vector. Although the causal time delay is already taken into account, the temporal kinematic information before the neuron fires actually has no causal relation with respect to the current spike. For the instantaneous kinematic encoding model, the optimum time delay has to be estimated to draw as much information as possible. The instantaneous motor cortical neural activity can be modelled as

\[ \lambda_t = f(\vec{k} \cdot \vec{x}_{t+lag}) \]  

\[ \text{spike}_t = \text{Poisson}(\lambda_t) \]

where \( \vec{x}_{t+lag} \) is the instantaneous kinematics vector defined as \([\vec{p}_x \, \vec{v}_x \, \vec{a}_x \, \vec{p}_y \, \vec{v}_y \, \vec{a}_y]_t^{\text{lag}} \) with 2-dimensional information of position, velocity and acceleration with causal time delay. \( \vec{k} \) is a linear filter, representing the preferred instantaneous direction in high-dimensional kinematics space. The weight estimation of the linear filter is based on the standard technique of spike-triggered regression.

\[ \vec{k} = \left( E[\vec{x}_{t+lag}^T \vec{x}_{t+lag}] + \sigma \right)^{-1} E_{\vec{x}_{t+lag} \mid \text{spike}_t}[\vec{x}_{t+lag}] \]  

Equation 3 represents the least square solution for the linear adaptive filter, where \( E[\vec{x}_{t+lag}^T \vec{x}_{t+lag}] \) gives the autocorrelation matrix \( R \) of the input vector considering causal time delay. \( \sigma \) is a regularization factor, which avoids ill-conditioning in the inverse. In the experiment, \( \sigma \) is chosen to maximize the linear filter performance. From a statistical perspective, \( E_{\vec{x}_{t+lag} \mid \text{spike}_t}[\vec{x}_{t+lag}] \) mimics the role of the cross-correlation vector \( P \) between the input and the binary spike train considering a causal time delay. Therefore, Equation 3 reduces to a conditional expectation of the binary spike train, i.e. this linear filter gives the spike-triggered average instantaneous kinematic vector \( E_{\vec{x}_{t+lag} \mid \text{spike}_t}[\vec{x}_{t+lag}] \) scaled by the decorrelated kinematic data \( E[\vec{x}_{t+lag}^T \vec{x}_{t+lag}] + \sigma \)^{-1}.

\[ \lambda_t \] is the instantaneous firing rate in an inhomogeneous Poisson spike generator. For the time interval selected for the spike analysis (i.e. the time interval valid for a Poisson assumption in the collected data, which has to be experimentally determined), a number is randomly drawn from a normalized uniform distribution (i.e. 0 to 1) and compared with the instantaneous conditional firing probability. If the number is smaller than the probability, then a spike is generated in this time interval.  

---

1 The error accumulation was experimentally verified in our decoding results. The detail of the decoding algorithm is out of the scope of this paper. Please refer to [Wang et al.2007a] if interested.
This modelling approach is therefore intrinsically stochastic, which carries implications (large variance) to on-line real time implementations.

$f$ is the nonlinear function estimated by an intuitive nonparametric technique [Chichilnisky 2001, Simoncelli et al. 2004] as the conditional probability density $p(spk \mid \tilde{k} \cdot \tilde{x})$ directed from the data. It is the fraction of the two kernel smoothed histograms of marginal $p(\tilde{k} \cdot \tilde{x})$ and joint distribution $p(spk, \tilde{k} \cdot \tilde{x})$. The histogram of the spike-triggered angle is smoothed by a Gaussian kernel according to Silverman’s rule [Silverman, 1981] and normalized to approximate the joint probability $p(spk, \tilde{k} \cdot \tilde{x})$, depicted as the solid red line in upper plot of Figure 2. In other words, the direction angle is accounted for in the histogram during the corresponding direction angle bin only when there is a spike. Then the conditional probability density $p(spk \mid \tilde{k} \cdot \tilde{x})$, depicted as the line in the bottom plot of Figure 2, is obtained by dividing the kernel-smoothed histogram of $p(spk, \tilde{k} \cdot \tilde{x})$ by the kernel-smoothed histogram of $\tilde{k} \cdot \tilde{x}$ (dotted line in the upper plot of Figure 2), which in fact implements Bayes rule,

$$p(spk = 1 \mid \tilde{k} \cdot \tilde{x}) = \frac{p(spk = 1, \tilde{k} \cdot \tilde{x})}{p(\tilde{k} \cdot \tilde{x})} \tag{4}$$

where $p(spk = 0 \mid \tilde{k} \cdot \tilde{x}) = 1 - p(spk = 1 \mid \tilde{k} \cdot \tilde{x})$. When $p(\tilde{k} \cdot \tilde{x})$ is 0, $p(spk = 1, \tilde{k} \cdot \tilde{x})$ is set to be 0.

![Marginal probability and joint probability](image1)

Fig. 2 The conditional probability density estimation

The peak in the conditional probability of Figure 2 is associated with the maximal firing probability, which is linked with specific values of the kinematic variables, and produces an increase in the firing rate of the neuron. Likewise, the region of low probability shows a deviation from the spontaneous firing rate for the neuron. These two portions of the curve (the most difficult to estimate well because they are at the tails of the distribution) are responsible for the modulation that is seen in
the rasters of the spike train data when observed along with the kinematic variables, and that are fundamental for BMI decoding performance.

2.2.2 Information theoretic Delay estimation

The causal time delay can also be estimated by information theoretical analysis. Here, we are interested in the optimum time lag, which extracts the most instantaneous kinematic information corresponding to the neural spike event. The well-established concept of mutual information [Reza 1994] as a metric for evaluating neuron instantaneous receptive properties is based on information theory and would capture much more of the neuronal response [Paninski et al 2004b, Wang et al. 2007b]. Define a tuned cell as a cell that extracts more information between the linear filtered kinematics and its spiking output. If a neuron is tuned to a preferred direction in high-dimensional space, the mutual information between the spike and the delayed linear filter kinematics vector is first drawn simply as the function of the time lag after a spike as in (5).

$$I_{(spk;\tilde{x}(lag))} = \sum_{k} \sum_{spk=0,1} \log_{2} \left( \frac{p(spk | k \cdot \tilde{x}(lag)) p(k \cdot \tilde{x}(lag))}{p(spk)} \right)$$

where $p(k \cdot \tilde{x}(lag))$ is the probabilistic density the linear filtered kinematics as a function of time lag, which can be easily estimated by Parzen window [Parzen 1962]. $p(spk)$ can be calculated simply as the percentage of the spike count during the entire spike train. $p(spk | k \cdot \tilde{x})$ is exactly the nonlinear function $f$ in LNP model.

The time delay with the highest mutual information is assigned as the optimum time lag for each neuron. The kinematics at the optimum time lag carries maximally the causal information of the neural spike. In the encoding stage, the 6-dimensional kinematic vectors are first synchronized at the optimum delay for each neuron, then input to the LNP tuning model to generate the estimated firing rates according to equation (1). To test the encoding ability of the instantaneous tuning model, the neuron firing rate is obtained by smoothing the real spike train with a Gaussian kernel. The correlation coefficient is then calculated between two firing rates to measure the quality of encoding.

3. Results

3.1 Information theoretic Delay estimation

As we mentioned in the previous section, the windowed kinematic vector is usually chosen as 300 msec before and 500 msec after the current neural spike, which already takes into account the causal time delay of the motor cortical neurons. We selected a possible delay range from 0 to 500ms after a neuron spikes to estimate the optimum time delay for our instantaneous tuning function. The regularization factor $\sigma$ in the spike-triggered average stage is experimentally set as $10^{-7}$, and the kernel size to smooth the histogram of probability density is set according to Silverman’s rule [Silverman, 1981]. For all 185 neurons, the mutual information as a function of time delay was obtained from 10,000 continuous samples (100 seconds) during movement. The time delay with highest mutual information was assigned as the best time lag for each neuron. Since neurons in M1 show more tuning information than other cortical areas, here we study 5 neurons that show the highest tuning: neurons 72, 77, 80, 99, and neuron 108. Figure 3 shows the mutual information as a function of time delay after spike occurrence. The best time lags are marked by a cross on each curve, and are 110ms, 170 ms, 170ms, 130 ms and 250ms respectively. It is interesting to observe that not all the neurons have the same time delay, although all of these neurons are in M1. During the analysis, different time delay is used for each neuron respectively. The average best time delay for all 185 neurons was 220.108ms, which is close to the results mentioned in the literature [Wu et al. 2006].
3.2. Instantaneous versus windowed tuning curves

The windowed encoding approach yields a widely accepted exponential increasing nonlinear function \( f \) after linear projection [Paninski et al. 2004a, Paninski et al. 2004b]. However for BMIs we are proposing an instantaneous and global (i.e. across kinematic variables) tuning estimation, therefore it is important to compare and evaluate the two tuning methodologies. For each neuron, we chose 7 different window sizes to filter the kinematic vector \( \{\tilde{p}_x, \tilde{v}_x, \tilde{a}_x, \tilde{p}_y, \tilde{v}_y, \tilde{a}_y\}_T^{t+\tau} \) and calculate the nonlinearity using the methods described in section 2. The biggest window size is 300 ms before and 500 ms after the current neural spike, noted as [-300, 500], which has been used in [Paninski et al. 2004b] for motor cortical neuron tuning analysis. Then each window shrinks 50ms at left and right extremes, such as [-250, 450], [-200, 400], ... until the smallest window [0, 200] ms. Figure 4 shows the nonlinearity of the 4 MI neurons estimated by windowed kinematics with 7 different window sizes, each plotted in different colours. The instantaneous nonlinear tuning with optimum delay is emphasized in a thick red line.
As we can observe from the figures, the tuning curves vary with different window sizes, particularly in the high tuning region. However, the middle part of the nonlinearity is very stable across all the window sizes, including the instantaneous estimation. Compared to the windowed tuning, the instantaneous model produces a smaller dynamic range of projected values (x-axis) because it directly works in the dynamics range of the kinematics without involving time-embedded information. We chose the Correlation Coefficient (CC) as the criterion to evaluate the similarity between the nonlinear tuning curves estimated from each windowed kinematics and the instantaneous one within the range specified by the instantaneous model. Seven histograms of correlation coefficients are shown in Figure 5, where the y-axis shows the percentage of neurons (out of 185) with a given CC. We can see that 98.92% of neurons have instantaneous tuning curves with a similarity over 0.9 compared to the one by window size [-300, 500] ms. More than half (58.38%) of the neurons have a similarity over 0.9 for the [-50, 250] ms window. However, less than half (41.62%) of the neurons have a similarity over 0.9 for the [0, 200] ms window because this window is not big enough to include the optimum causal delay, which is on average 220 ms. Since the summation for the same window size (color bar) is 100%, the similarity of the less similar neurons (CC<0.9) is distributed.
across other CC bins. Also notice that from the windowed methods the one with the smallest window, when it includes the optimum time delay, is the closest to the instantaneous estimated tuning. The similarity amongst windowed and instantaneous methods is rather surprising, and builds confidence that in spite of its simplicity in computation it is quantifying appropriately neural tuning properties.

Fig. 5 Correlation coefficient between the nonlinearity calculated from windowed kinematics and the instantaneous kinematics with optimum delay

One possible reason for the differences at both extremes of the tuning curves is insufficient data to provide an accurate estimation at both extremes, in particular because of the division in (4). Recall that this is actually the important part of the tuning curve for BMIs because it is in this portion that the neuron firing shows modulation with the kinematic variable. In particular, neurons 80 and 99 (as many others) show a large mismatch at the high firing rate level (right end of the curve). Both neurons demonstrate a lower firing probability in the instantaneous curve compared to the windowed curves. Neuron 80 also shows a saddle-like behavior very different from the exponential increase. Therefore these behaviors need to be further investigated.

3.2.3. Instantaneous versus windowed encoding

Since the ultimate goal of the tuning analysis is to transform spike timing information into the kinematic space, here we compare both tuning methods in our experimental data set described in section 2.2. Here we select neuron 80 and neuron 99 to compare the encoding ability between the windowed and the instantaneous tuning model with the real kinematics signals (Figure 6(a) and 6(b)). From previous studies, these 2 neurons are known to be among the most sensitive neurons for BMI modeling [Sanchez et al. 2003, Wang et al. 2007b], and they are also amongst the ones that show the larger mismatch at the high firing probability range (right extreme end of Fig 4).

In each plot, the pink bars in the first and second rows represent the neural spike train. The red dash line superimposed on the spike train is the firing rate estimation by kernel smoothing. In the top panel,
the blue solid line superimposed on the spike train is the estimated firing rate by instantaneous tuning, while in the second panel, the green solid line superimposed on the spike train is the estimated firing rate by windowed tuning with 300 ms before and 500 ms after the current neuron firing. To check the animal’s behavior simultaneously with the neural activity, the third and fourth panels show the re-scaled 2D position and velocity (blue for x, green for y) after synchronization at the optimum delay.

We can clearly observe that, for both neurons (Fig 6 (a) and (b)), the instantaneous model gives a smoother estimated firing rate than the noisy estimation by the windowed model. We found that the linear filter outputs in the windowed model are very noisy, because it is a projection of the high dimensional time-embedded kinematic vector, which increases the range of the independent variable and so creates larger variability in the spike rates. Moreover, the over-estimation at high firing rate of the nonlinearity curve leads to the extraneous large peaks on the green line. As can be expected, since the tuning is higher, there will be more spikes and so the intensity function estimation is very high and noisier as seen in the green curve.

It is also very interesting to notice that after causal alignment both neurons demonstrate clear time alignment (negative correlation) between the hand velocity trajectory and the peaks of firings, which reinforces the evidence for neural kinematic encoding.

![Fig. 6(a) Comparison of Encoding results by instantaneous modelling and windowed modelling for neuron 99](image-url)
To quantify the encoding comparisons, the correlation coefficient between the neuron’s firing rate and the estimated firing rates from the windowed and instantaneous models. The kernel size smoothes the spike train to enable the estimation of CC but it will affect the results of the similarity measure. Figure 7(a) and (b) shows results comparing the CC for the same 2 neurons vs. different kernel sizes. Correlation coefficients for the instantaneous model are always greater than the ones by windowed model across kernel sizes. Here we choose to display the kernel size that maximizes the similarity. For neuron 99, the correlation coefficient between the instantaneous model and the firing rate is 0.6049, which is greater than 0.4964 for the windowed model. For neuron 80, the correlation coefficient between the estimated firing rate with the instantaneous model and the firing rate from real spike train is 0.6393, which is greater than 0.5841 given by the windowed model. Therefore, the instantaneous model shows better encoding ability.
modelling for neuron 99 across kernel size  
modelling for neuron 80 across kernel size

4. Conclusions and future work

The traditional criterion of estimating tuning depth from windows of data does not seem the most appropriate in the design of BMIs using sequential estimation algorithms on spike trains. Here we present instead an information theoretical tuning analysis of instantaneous neural encoding properties that relate the instantaneous value of the kinematic vector to neural spiking. The proposed methodology is still based on the LNP model, and an information theoretic formulation provides a more detailed perspective when compared with the conventional tuning curve because it statistically quantifies the amount of information between the kinematic vectors triggered off by the spike train. As a direct consequence, it can estimate the optimum time delay between motor cortex neurons and behavior caused by the propagation effects of signals in the motor and peripheral nervous system.

The similarities and differences between the windowed and instantaneously evaluated tuning functions were also analyzed. We conclude that the instantaneous tuning curves for most of the neurons show over 0.9 correlation coefficients in the central region of the tuning curve, which unfortunately is not the most important for BMI studies. There are marked differences in the high tuning region of the curves, both in the dynamic range and in the estimated value. The windowed model works on a time-embedded vector, which spreads the linear output $\vec{k} \cdot \vec{x}$ to a wider range. Since the $pdf$ integral is always 1, the windowed model flattens the marginal distribution of $p(\vec{k} \cdot \vec{x})$. In the time segment when the neuron keeps firing, the overlapping windows make the linear filter output $\vec{k} \cdot \vec{x}$ change slowly. It results in more spike-triggered samples in the small neighbourhood of $\vec{k} \cdot \vec{x}$. Therefore, the estimation on the joint distribution $p(spk, \vec{k} \cdot \vec{x})$ becomes higher. Both consequences contribute to the overestimation of tuning at the high firing rate part of the windowed nonlinear curve.

The instantaneous model works directly in the dynamic range of the kinematics that is sensitive only to the corresponding neuron spike timings. It estimates more accurately the firing probability without distortions from the temporal neighbourhood information. However, we create a vector with all of the kinematics (position, velocity, acceleration) to estimate better (i.e. to obtain more sensitivity) the tuning from the data. This has the potential to mix tuning information for the different kinematics variables and different directions if they are not exactly the same. When the different kinematic variables display different sensitivities in the input space, after projection by the weight filter direction they will peak at different values of $\vec{k} \cdot \vec{x}$ in the nonlinear curve, which then results in the saddle-like feature observed in Fig. 10. The other potential shortcoming is that less data is used, so the variability may be higher. However, at this time one still does not know which tuning curve provides a better estimate for the instantaneous tuning model required in the encoding and decoding stages of BMIs. Ultimately, the instantaneous model can produce equivalent or better encoding results compared to existing techniques. This outcome builds confidence to directly implement the instantaneous tuning function into the future online decoding work for Brain-Machine Interfaces.

Acknowledgments

This work was supported by NSF grant 0422718 and #CNS-0540304 and the Children’s Miracle Network. The authors would like to thank Dr. Miguel Nicolelis for the use of the monkey data collected in his laboratory at Duke University.

References

DeAngelis G C, Ohzawa I and Freeman R D 1993 The spatiotemporal organization of simple cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation Journal of Neurophysiology 69 1118-1135
Fritsch G and Hitzig E 1870 Ueber dir elektrische Erregbarkeit des Grosshirns Arch. Anat. Physiol. Lpz. 37 300-332
Jones J P and Palmer L A 1987 The two-dimensional spatial structure of simple receptive fields in cat striate cortex Journal of Neurophysiology 58(6) 1187-1211
Kim S P 2005 Design and analysis of optimal encoding models for brain machine interfaces PhD. Dissertation University of Florida
Leyton A S F and Sherrington C S 1917 Observations and excitable cortex of the chimpanzee, orangutan and gorilla Q. J. Exp. Physiol. 11 135-222
McLean J and Palmer L A 1989 Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat Vision Research 29 675-679
Moran D W and Schwartz A B 1999 Motor cortical representation of speed and direction during reaching The Journal of Neurophysiology 82(5) 2676-2692
Nicolelis M A L, Ghazanfar A A, Faggin B, Votaw S, and Oliveira L M O 1997 Reconstructing the engram simultaneous, multiple site, many single neuron recordings Neuron 18 529-537
Reid R C and Alonso J M 1995 Specificity of monosynaptic connections from thalamus to visual cortex Nature 378(6554) 281-284
Schwartz A B, Kettner E, and Georgopoulos A P 1988 Primate motor cortex and free arm movements to visual targets in three-dimensional space. I. Relations between single cell discharge and direction of movement *J. Neurosci.* 8 2913–2927