

# Template-Based Spike Pattern Identification With Linear Convolution and Dynamic Time Warping

Zhiyi Chi, Wei Wu, Zach Haga, Nicholas G. Hatsopoulos and Daniel Margoliash  
*J Neurophysiol* 97:1221-1235, 2007. First published Nov 15, 2006; doi:10.1152/jn.00448.2006

**You might find this additional information useful...**

---

This article cites 40 articles, 23 of which you can access free at:

<http://jn.physiology.org/cgi/content/full/97/2/1221#BIBL>

Updated information and services including high-resolution figures, can be found at:

<http://jn.physiology.org/cgi/content/full/97/2/1221>

Additional material and information about *Journal of Neurophysiology* can be found at:

<http://www.the-aps.org/publications/jn>

---

This information is current as of February 22, 2007 .

# Template-Based Spike Pattern Identification With Linear Convolution and Dynamic Time Warping

Zhiyi Chi,<sup>1,4</sup> Wei Wu,<sup>2</sup> Zach Haga,<sup>3</sup> Nicholas G. Hatsopoulos,<sup>3</sup> and Daniel Margoliash<sup>3</sup>

<sup>1</sup>Department of Statistics, University of Connecticut, Storrs, Connecticut; <sup>2</sup>Department of Statistics, Florida State University, Tallahassee, Florida; and <sup>3</sup>Department of Organismal Biology and <sup>4</sup>Anatomy, University of Chicago, Chicago, Illinois

Submitted 28 April 2006; accepted in final form 15 November 2006

**Chi Z, Wu W, Haga Z, Hatsopoulos NG, Margoliash D.** Template-based spike pattern identification with linear convolution and dynamic time warping. *J Neurophysiol* 97: 1221–1235, 2007. First published November 15, 2006; doi:10.1152/jn.00448.2006. Pattern identification for spiking activity, which is central to neurophysiological analysis, is complicated by variability in spiking at multiple timescales. Incorporating likelihood tests on the variability at two timescales, we developed an approach to identifying segments from continuous neurophysiological recordings that match preselected spike “templates.” At smaller timescales, each component of the preselected pattern is represented by a linear filter. Local scores to measure the similarities between short data segments and the pattern components are computed as filter responses. At larger timescales, overall scores to measure the similarities between relatively long data segments and the entire pattern are computed by dynamic time warping, which combines the local similarity scores associated with the pattern components, optimizing over a range of intercomponent time intervals. Occurrences of the pattern are identified by local peaks in the overall similarity scores. This approach is developed for point process representations and binary representations of spiking activity, both deriving from a single underlying statistical model. Point process representations are suitable for highly reliable single-unit responses, whereas binary representations are preferred for more variable single-unit responses and multiunit responses. Testing with single units recorded from individual electrodes within the robust nucleus of the arcopallium of zebra finches and with recordings from an array placed within the motor cortex of macaque monkeys demonstrates that the approach can identify occurrences of specified patterns with good time precision in a broad range of neurophysiological data.

## INTRODUCTION

Neurons exhibit patterned activity at multiple scales of time (Reich et al. 2000, 2001; Zugaro et al. 2004). Analysis of such data, often obtained as extracellular recordings, is central to the goals of neurophysiology. A subset of this general problem is identification of spiking activity patterns in one record that match previously identified patterns in another record. For example, the analysis of sleep mechanisms of learning and memory has in part involved comparison of ongoing discharge during sleep with discharges recorded during wakeful behavior (Buzsáki 1989). Based on the identified patterns, some inferences can be made on the information processing of the brain during sleep (Buzsáki 1998; Dave and Margoliash 2000; Hahnloser et al. 2006; Louie and Wilson 2001, 2002; Nádasdy 2000; Skaggs and McNaughton 1996; Wilson and McNaughton 1994). This is a difficult problem because of the variability of neuronal activity in sleep relative to wakeful behavior and

because there is no time referent during sleep. As another example, from a neural decoding perspective it is of interest to estimate the times of certain landmark events during a course of behavior based on the associated spiking activity (Chi and Margoliash 2001). The estimation of behavioral timing can be useful when it is combined with procedures designed to estimate spatial trajectories of motor output (Georgopoulos et al. 1986; Hatsopoulos et al. 2004; Smith and Brown 2003; Wu et al. 2004, 2006); however, it appears that this type of estimation has not received much attention in the literature. We show herein that pattern identification for spiking activity can provide a good estimate for behavioral timing.

Template matching has been the basis of many pattern-identification procedures (Abeles et al. 1988, 1993; Chi et al. 2003; Dave and Margoliash 2000; Louie and Wilson 2001; Nádasdy et al. 1999). In these procedures, a template is constructed based on exemplar spike trains that exhibit a certain pattern of interest. During data analysis, each segment of the data is scored in terms of its similarity to the template. Segments with high similarity scores are interpreted as exhibiting the pattern of interest and are thus extracted for further analysis.

In most of the current procedures, the template is treated as nearly time locked to stimuli or behavioral events. Except for some local variability, such as random perturbation in spike timing, global variability is expressed as uniform timescaling, which multiplies all the interspike intervals (ISIs) by a single factor (Chi et al. 2003; Louie and Wilson 2001; Nádasdy et al. 1999). However, in some systems spiking activity can exhibit “time warping,” i.e., nonuniform time compression or expansion on a relatively large timescale (Dave and Margoliash 2000; Nádasdy et al. 1999). As far as we know, the procedure introduced by Victor and Purpura (1996) is the only one that explicitly addresses time warping. The procedure was developed to analyze the precision of temporal coding during specific short time intervals of sensory stimulation. In the procedure, the ISIs are assumed to vary independently and similarity scores are computed with dynamic time warping (DTW). Such scores are known as “text edit distances” in the field of computer science (Wagner and Fischer 1974). Similar procedures based on text edit distances are extremely useful in bioinformatics (Aach and Church 2001; Bishop and Thompson 1986; Needleman and Wunsch 1970; Waterman et al. 1987). Sequences of discrete behaviors such as syllables in birdsongs are also occasionally analyzed using the text edit distance

Address for reprint requests and other correspondence: Z. Chi, 215 Glenbrook Rd., U-4120, Storrs, CT 06269 (E-mail: zchi@merlot.stat.uconn.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

(Margoliash et al. 1991; Sankoff and Kruskal 1983). DTW is also commonly used in speech recognition (Huang and Gray 1988; O'Shaughnessy 1999; Rabiner and Juang 1993), bird-song recognition (Anderson et al. 1996; Ito and Mori 1999; Kogan and Margoliash 1998), and other applications.

The variability of spiking activity at local versus relatively large timescales can be roughly described as follows (Fig. 1A). Think of the occurrences of a pattern of interest as "texts." The "codewords" of the texts are short intervals with spikes. The definition of *codeword* should be problem dependent. For different systems, a codeword may contain a single spike, multiple spikes, or a high-frequency burst. Within the codewords, the detectable variability can be reasonably characterized as random local perturbation in spike timing and spike count. Between the codewords, on the other hand, the gaps often have quite visible nonuniform variability, resulting in time warping. The gaps need not be activity free. Instead, they can contain "noise" spikes that do not repeat in different occurrences of the pattern.

Because the perturbation within codewords and the time warping occur at different timescales, they should be treated differently; trying to fit them into a single statistical model is likely to end up with mischaracterization of one or both. From this perspective, the template-based approach proposed in this report consists of two steps: 1) applying linear convolution to compare short data segments with individual codewords of the template to yield "local" similarity scores; and 2) applying

DTW to combine the local similarity scores to attain optimal "global" similarity scores. Data segments with patterns similar to the template are identified by above-threshold local peaks in the global similarity scores.

The local comparison should be based on specific representations of spiking activity. We shall consider two types of representations. The point process representation registers spike times continuously. As a result, the neural activity is characterized as processes of points in time. The binary representation divides the neural activity into small time bins, and assigns 1 to a bin if there is at least one spike in it and 0 otherwise. As a result, the neural activity is characterized as sequences of 0 and 1 s. Although the local comparison algorithms for these two representations look quite different, they derive from nearly identical arguments based on likelihood tests.

By using DTW, the template-based approach automatically "parses" each identified data segment into codewords and gaps in between, thus allowing a more detailed interpretation of the identification results (Fig. 1A). Such parsing can yield additional insight. If the spiking activity pattern is preselected in association with a certain behavior, and each of its codewords is reliably associated with a landmark event during the behavior, then the parsing provides a way to infer from the neural activity when the landmark events occur within the overall behavior. Presently, the dominant procedures for inference of the time course of behavior are based on regression or state-space models (Brockwell et al. 2004; Carmena et al. 2003; Serruya et al. 2002; Smith and Brown 2003; Wu et al. 2004, 2006). They are designed to estimate spatial trajectories of motor output with minimized average error over time. The estimation of the timing of events during motor output can therefore provide a useful new perspective to the decoding of neuronal activity.

In what follows, under METHODS, we describe the identification procedures and present pseudocode for their implementations. The procedure for the point process representation was implemented in C++, and that for the binary representation in MATLAB (The MathWorks, Natick, MA). Under RESULTS, we illustrate the utility of the procedures with two brief neurobiological examples. The procedure for the point process representation is applied to a single-unit (SU) recording collected during sleep from the nucleus robustus arcopallium (RA) of zebra finches (Dave and Margoliash 2000). The procedure for the binary representation is applied to multielectrode recordings from the primary motor cortex (MI) of a behaving macaque monkey during a target pursuit task (Hatsopoulos et al. 2004; Serruya et al. 2002). In this example, the times when a target was reached and the following target appeared in each successful trial can be estimated with error  $<0.2$  s. This is somewhat surprising in view of the relatively high level of spike-timing variability in MI. For each procedure, we conduct a sensitivity analysis, demonstrating that the procedures are robust.

## METHODS

The neuronal recordings analyzed in RESULTS were obtained from adult male zebra finches (*Taeniopygia guttata*) and *Macaca mulatta* monkeys. The experiment procedures were conducted under protocols approved by the University of Chicago IACUC for birds (Margoliash) and monkeys (Hatsopoulos).

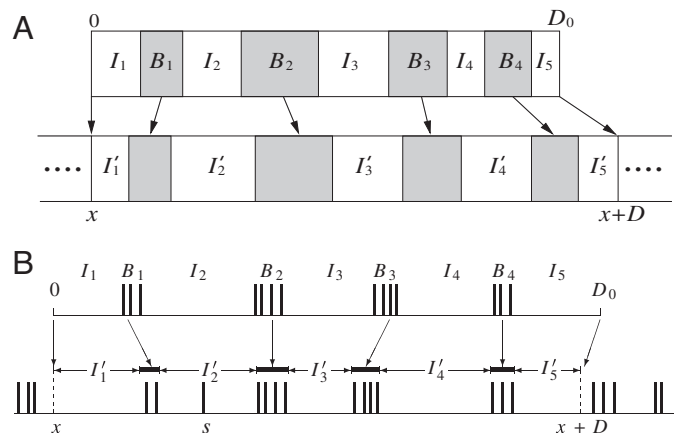


FIG. 1. A schematic of template matching. *A, top*: a sketch of a template registered in  $[0, D_0]$ . Template may be a single-unit (SU) spike train, an array of firing rates of multiple single units, or other suitable representations of spiking activity. Shaded blocks  $B_1, \dots, B_4$  represent "codewords" of the template that are almost "rigid" and can be perturbed only locally, whereas the blank blocks  $I_1, \dots, I_5$  represent intervals that can be compressed or expanded in time. *Bottom*: matching the template with a data segment at time  $x$ . Arrows define how different parts of the template and those of the segment are matched. Onset of the template is mapped to the onset  $x$  of the data segment. By compressing or expanding  $I_i$ , the codewords are aligned and compared with the data in the shaded blocks indicated by the arrows. Overall similarity score between the template and the data segment is the sum of the similarity scores between the codewords and the corresponding data intervals, minus a weighted sum of the modifications on the intervals  $I_i$ . Depending on problems, the spiking activity in the intervals  $I_i$  of the data are either ignored or treated as noise and included in the overall similarity score as a minus term. *B*: template matching for single-unit spike trains, as for the nucleus robustus arcopallium (RA) data. *Top*: each  $B_i$  is a burst in the template and  $I_i$  is an interburst interval (IBI). *Bottom*: bursts in the data segment are matched with the template bursts. Spike  $s$  in the data segment does not correspond to any burst in the template and is counted as a "noise" spike. Amount of noise in  $I_i$  is evaluated by the number of spikes in them.

Because of the nature of the recording technology and cortical activity, the levels of accuracy of these two recordings were very different. First, in terms of signal-to-noise ratio (SNR), the accuracy of separating action potentials from background was much higher for the recordings from RA. The SNRs of spikes from SU recordings in the RA of zebra finches are generally  $>10$ , whereas those from the array recordings in macaque monkeys can be as low as 3. The difference in SNRs result from the different experimental designs and thus different recording technologies are used: movable but small numbers (one to four) of electrodes in the bird versus a fixed array of large numbers ( $>100$ ) of electrodes in the monkey. Second, in terms of neurophysiological characteristics, the spike-timing variability of zebra finch RA units is much lower as well (Yu and Margoliash 1996; cf. Hatsopoulos et al. 2004). The differences are likely to influence the performance of pattern identification. The sampling rate of recording is 20 kHz for the RA data and 30 kHz for the MI data.

### Part one: point process representation

The procedure described in this part is suitable when 1) the pattern of interest consists of well-defined spike bursts; 2) across occurrences of the pattern, the bursts exhibit low variability; and 3) the data being analyzed constitute a single spike train, consisting of time registry of spikes accurately estimated from the SU recording. Under these conditions, an exemplar spike train expressing the pattern may be used as a template. Under RESULTS, more detail will be given on how templates are selected for SU activity in RA.

The procedure directly compares the spike times. Figure 1B illustrates how to compare a template with data registered around a time location. In general, the template need not start or end with a spike. Together with its bursts, the intervals at the beginning and at the end of the template and those between the bursts are all parts of the template and will be referred to as *interburst intervals* (IBIs). The bursts are used as the codewords of the template pattern. The IBIs can be thought of as independent “springs,” and the bursts as relatively incompressible “blocks” interconnected by the springs. By compressing or stretching the IBIs, the template bursts are shifted accordingly and compared with the data registered in the corresponding time intervals. Given a set of modified IBIs, the score of similarity between the warped template and the data segment aligned with it is a sum of the similarity scores associated with the bursts, minus a weighted sum of the amount of changes in the IBIs and the number of spikes in the data segment that do not match any template spikes. Optimizing the weighted sum over a range of modified IBIs then yields the optimal score at the time location. We next consider appropriate similarity scores and their computation.

**TEMPLATE.** A template spike train has to be registered in an interval  $[0, D]$  (Fig. 1B). If the template has  $n$  bursts, then in general it has  $n + 1$  IBIs, including the interval between its onset ( $t = 0$ ) and first spike, and the interval between its last spike and offset ( $t = D$ ). A subset of template spikes is a burst if all its ISIs are less than a specified threshold. Fix parameter  $\Delta > 0$ , which controls the precision for comparing spike times. Let  $H_{n+1} = D$ ,  $T_0 = 0$ , and for  $1 \leq i \leq n$

$$H_i = \text{the } i\text{th burst's first spike} - \Delta, \quad T_i = \text{the } i\text{th burst's last spike} + \Delta$$

The  $i$ th burst interval is defined as  $[H_i, T_i]$  with duration  $B_i = T_i - H_i$ ; the  $i$ th IBI is defined as  $[T_{i-1}, H_i]$  with duration  $I_i = H_i - T_{i-1}$  ( $H$  stands for “head”;  $T$  for “tail”).

**LOCAL SCORES FOR TEMPLATE BURSTS.** First, we fix a kernel function  $K(x)$ . This will be used to measure the temporal discrepancy between a data spike and a template spike. Some commonly used kernel functions defined on  $[-1, 1]$  are constant 1 (“square” kernel),  $1 - |x|$  (“triangular”),  $1 - x^2$  (“Epanechnikov”), and  $(1 - x^2)^2$  (“biweight”). To evaluate the similarity between the  $i$ th template burst

and a data burst  $s_1, s_2, \dots$  in  $[x, x + B_i]$ , the idea is to translate both to  $[0, B_i]$  so they can be compared directly. If the  $i$ th template burst consists of spikes  $t_1, t_2, \dots$ , then the similarity score between the two bursts is

$$F_i(x) = \sum_j \Phi_i(s_j - x) \quad \text{with} \quad \Phi_i(s) = (1 + \nu) \max_j K\left(\frac{s - (t_j - H_i)}{\Delta}\right) - \nu \quad (1)$$

with  $\nu \geq 0$  the maximum absolute negative contribution a nonmatching data spike can make to the local score. From Chi et al. (2003),  $F_i(x)$  can be computed as a linear convolution of the entire data spike train with  $\Phi_i$ .

**GLOBAL SCORE FOR THE ENTIRE TEMPLATE.** Denote by  $N(a, b)$  the number of data spikes between  $a$  and  $b$ . For the  $i$ th IBI, let  $G_i(x) \geq 0$  be a cost function for its warping. Suppose the warped template has its IBIs changed by  $v_1, \dots, v_{n+1}$ . Denote

$$V_i^k = \begin{cases} v_1 + \dots + v_k & \text{if } i \geq k \\ 0 & \text{otherwise} \end{cases}$$

Then the total similarity score between the warped template and the data at  $x$  is

$$l(x, v_1, \dots, v_{n+1}) = \sum_{i=1}^n F_i(x + H_i + V_i^i) - \sum_{i=1}^{n+1} G_i(v_i) - \nu \sum_{i=1}^{n+1} N(x + T_{i-1} + V_i^{i-1}, x + H_i + V_i^i) \quad (2)$$

Equation 2 is derived in the appendix and can be explained by Fig. 1B. By changing the IBIs, the  $i$ th template burst is aligned to the data in  $[x + H_i + V_i^i, x + T_i + V_i^i]$ , yielding the first term on the right-hand side of Eq. 2. The second term penalizes the changes on the template IBIs. In the warped template, the IBIs are  $[x + T_{i-1} + V_i^{i-1}, x + H_i + V_i^i]$ ,  $1 \leq i \leq n + 1$ . Data spikes in these intervals are counted as “noise” and reduce the similarity. This gives rise to the third term in Eq. 3.

By the probabilistic interpretation given in the APPENDIX, the overall similarity score assigned to the time location  $x$  is defined as

$$L(x) = \max_{v_1, \dots, v_{n+1}} l(x, v_1, \dots, v_{n+1}) \quad (3)$$

In general, there can be multiple solutions to the optimization in Eq. 3, each one being a set of changes in the IBIs. Let  $\hat{v}_1(x), \dots, \hat{v}_{n+1}(x)$  be a solution. By Fig. 1B, we can establish a one-to-one correspondence between the intervals  $I_1, \dots, I_{n+1}$  in the template and  $I'_1, \dots, I'_{n+1}$  in the data segment and identify the latter as a set of optimal matching IBIs. Then we can identify the intervals  $B'_1, \dots, B'_n$  between  $I'_1, \dots, I'_{n+1}$  as a set of optimal matching burst intervals. The interlacing  $I'_1, B'_1, I'_2, \dots, B'_n, I'_{n+1}$  form a partition of the data segment between  $x$  and the right endpoint of  $I'_{n+1}$ , which is an optimal match to the template that one can possibly find at time  $x$  under the similarity score. The optimal matching IBIs  $I'_i$  and burst intervals  $B'_i$  are computed as follows

$$I'_i = [x + T_{i-1} + \hat{V}_1^{i-1}(x), x + H_i + \hat{V}_i^i(x)] \quad 1 \leq i \leq n + 1$$

$$B'_j = [x + H_j + \hat{V}_j^j(x), x + T_j + \hat{V}_j^j(x)] \quad 1 \leq j \leq n \quad (4)$$

Because of the one-to-one correspondence between their subintervals, we can inspect in more detail how similar the optimal matching data segment and the template are.

DTW FOR THE GLOBAL COMPARISON. To compute  $L(x)$  and  $\hat{v}_1(x), \dots, \hat{v}_{n+1}(x)$ , let

$$l_k(x, v_k, \dots, v_{n+1}) = \sum_{i=k}^n F_i(x + H_i + V_i^k) - \sum_{i=k}^{n+1} G_i(v_i) - \nu \sum_{i=k}^{n+1} N(x + T_{i-1} + V_{i-1}^{i-1}, x + H_i + V_i^i)$$

$$L_k(x) = \max_{v_k, \dots, v_{n+1}} l_k(x, v_k, \dots, v_{n+1})$$

$$M_k(x, v) = F_k(x + H_k + v) - G_k(v) - \nu N(x + T_{k-1}, x + H_k + v)$$

Clearly,  $l(x, v_1, \dots, v_{n+1}) = l_1(x, v_1, \dots, v_{n+1})$  and thus  $L(x) = L_1(x)$ . Because

$$l_k(x, v_k, \dots, v_{n+1}) = M_k(x, v_k) + l_{k+1}(x + v_k, v_{k+1}, \dots, v_{n+1})$$

it follows that  $L_k(x) = \max_v [M_k(x, v) + L_{k+1}(x + v)]$ . Therefore  $L(x)$  and  $\hat{v}_1(x), \dots, \hat{v}_{n+1}(x)$  can be computed by DTW. Table 1 summarizes the procedure.

FINDING MATCHING SEGMENTS. Data segments that potentially match the template can be identified by above-threshold local maximum points in  $L(x)$ . Specifically, if  $x$  satisfies the following criteria, then the data in  $[x, x + H_{n+1} + \hat{V}_{n+1}^{n+1}(x)]$  are extracted.

- 1)  $L(x) \geq \theta$ , where  $\theta$  is a given threshold;
- 2)  $L(x) = \max \{L(y) : |y - x| \leq r\}$ , with  $r$  the radius for local comparison; and
- 3)  $L(x) > L(y)$  for at least one  $y$  with  $|y - x| \leq r$ . This is to make sure that  $L$  is not a constant in  $[x - r, x + r]$ .

A reasonable choice for  $r$  is  $D$ , the duration of the template. When some of the extracted segments overlap in time, a disjoint subset of the segments may be selected.

The selection of the threshold value  $\theta$  in general is a difficult problem. One important reason is that the stochastic properties of the data can be difficult to grasp. In RESULTS we will describe an ad hoc method for the RA activity that tries to simulate the noise in the data. A different ad hoc method was used for the data collected from the MI of monkey, which will also be described in RESULTS.

ADAPTIVE SELECTION OF PARAMETER VALUES. The parameters  $\Delta$  and  $\nu$  can be set based on estimation. For  $\Delta$ , we first compute the mean value  $d$  of the template ISIs, which can be regarded as the minimum resolution to distinguish template spikes. If the square

TABLE 1. DTW for the point process representation: general form

1 Compute

$$\begin{aligned} v_{n+1}^*(x) &= \arg\max_v M_{n+1}(x, v) \\ L_{n+1}(x) &= M_{n+1}(x, v_{n+1}^*(x)) \end{aligned}$$

$$\text{with } M_{n+1}(x, v) = -G_{n+1}(v) - \nu N(x + T_n, x + H_{n+1} + v)$$

2 For  $i = n$  down to 1,

$$\begin{aligned} v_i^*(x) &= \arg\max_v [M_i(x, v) + L_{i+1}(x + v)] \\ L_i(x) &= M_i(x, v_i^*(x)) + L_{i+1}(x + v_i^*(x)) \end{aligned}$$

$$\text{with } M_i(x, v) = F_i(x + H_i + v) - G_i(v) - \nu N(x + T_{i-1}, x + H_i + v)$$

3 Let  $L(x) = L_1(x)$ , and  $\hat{v}_1(x) = v_1^*(x)$

4 For  $i=2, \dots, n+1$ ,

$$\hat{v}_i(x) = v_i^* \left( x + \sum_{k=1}^{i-1} \hat{v}_k(x) \right)$$

5 Return  $L(x)$  and  $\hat{v}_1(x), \dots, \hat{v}_{n+1}(x)$

kernel is used in Eq. 1 to compute the local scores, then set  $\Delta = d/2$ . To use a different kernel  $K$ ,  $\Delta$  has to be adjusted. The reason is that except for  $x = 0$ ,  $K(x)$  is strictly less than the square kernel. Without adjusting  $\Delta$ , the kernel puts less total weight to data spikes that match the template spikes, resulting in artificially lower similarity scores. We set

$$\Delta = c_K d/2 \quad (5)$$

where  $c_K > 1$  such that the area under  $K(x/c_K)$  is 2. The value of  $c_K$  is 2 for the triangular kernel, 1.5 for the Epanechnikov kernel, and 1.875 for the biweight kernel.

From Eq. A1 in the APPENDIX,  $\nu \approx \log(\lambda_0/\lambda)/\log(\lambda'/\lambda_0)$ , where  $\lambda'$  is the firing rate within template bursts,  $\lambda$  is the firing rate within template IBI, and  $\lambda_0$  is the mean firing rate in nonmatching spike trains. Because for a Poisson process the density is the reciprocal of the mean ISI, we set the value of  $\nu$  based on the approximation

$$\nu \approx \frac{\log(d/d_0)}{\log(d/d')} \quad (6)$$

where  $d'$ ,  $d$ , and  $d_0$  are the mean ISIs in template bursts, template IBIs, and the entire data spike train, respectively. In estimating  $d$ , if a template IBI has no spike, then its duration substitutes for a sample ISI. Whereas  $d'$  and  $d$  are fixed once templates are selected,  $d_0$  is dependent on data. Therefore the selection of  $\nu$  is adaptive.

NUMERICAL ISSUES. In data analysis,  $L(x)$  is computed on a discrete time grid  $jd$ , with  $j = 0, \pm 1, \dots$ , and  $d > 0$  a fixed step size. Also, the warping of each IBI has to be bounded. Table 2 contains a version of the algorithm that takes into account these constraints. It requires that  $F_i(x_0 + jd)$  be computed and stored for a given  $x_0$ . We apply the subroutine below to compute  $F_i(x_0 + jd)$ . Let  $\lceil a \rceil$  denote the smallest integer no less than  $a$ , and  $\lfloor a \rfloor$  the largest integer no greater than  $a$ .

- 1) Let  $F_i(x_0 + jd) = 0$  for all  $j$ .
- 2) For each data spike  $s$  and  $j = j_0, \dots, j_1$ , increase  $F_i(x_0 + jd)$  by  $\phi_i(s - x_0 - jd)$ , where  $j_0 = \lceil (s - x_0 - \Delta)/d \rceil$  and  $j_1 = \lfloor (s - x_0)/d \rfloor$ .
- 3) Return all  $F_i(x_0 + jd)$ .

### Part two: binary representation

The procedure described in this part is suitable when 1) the spiking activity is recorded from multiple single units and exhibits significant amount of variability in spike timing across trials and/or 2) the neural signals have lower SNRs compared with SU recordings; typically, the spike trains are obtained from multielectrode recordings by spike sorting. Under these conditions, the spike train of a single unit typically cannot represent the variability in a pattern of interest and therefore it alone cannot be directly used as a template for pattern identification. Instead, we shall construct a template as a linear filter that is made up of log-likelihood ratios estimated from sample spike trains of all the recorded single units. The statistical model involved in the construction is an integral part of the pattern identification procedure and is described below.

BASIC SETUP. Suppose the spiking activity pattern we are interested in is associated with a sequence of  $n$  landmark events during the course of a certain type of behavior. For example, if a behavioral task requires a monkey to reach  $n$  targets in sequence, then for  $i = 1, \dots, n$ , the  $i$ th event may be defined as the cursor reaching the  $i$ th target (see more detail in RESULTS). The  $n - 1$  intervals between the times of consecutive events during an occurrence of the event sequence will be referred to as *interevent intervals* (IEIs).

We suppose that the multielectrode spike trains are recorded from the same set of  $C$  units and each data spike has been associated with one of the units by spike sorting. Fix parameter  $\Delta > 0$  as the duration of time bin. For  $c = 1, \dots, C$ , the binary representation for the

TABLE 2. A discrete version of the DTW procedure in Table 1 that computes  $L(td)$ , with  $t$  an integer and  $d > 0$  a fixed step size

1 Compute

$$I_{n+1}^*(t) = \operatorname{argmax}_{-A_{n+1} \leq k \leq B_{n+1}} W_{n+1}(t, k)$$

$$Q_{n+1}(t) = W_{n+1}(t, I_{n+1}^*(t)), \quad t = 0, 1, \dots$$

with  $W_{n+1}(t, k) = -G_{n+1}(kd) - \nu N(td + T_n, (t + k)d + H_{n+1})$

2 For  $i = n$  down to 1,

$$I_i^*(t) = \operatorname{argmax}_{-A_i \leq k \leq B_i} [W_i(t, k) + Q_{i+1}(t + k)]$$

$$Q_i(t) = W_i(t, I_i^*(t)) + Q_{i+1}(t + I_i^*(t)), \quad t = 0, 1, \dots$$

with

$$W_i(t, k) = F_i((t + k)d + H_i) - G_i(kd) - \nu N(td + T_{i-1}, (t + k)d + H_i)$$

3 Let  $L(td) = Q_1(t)$ , and  $\hat{I}_1(t) = I_1^*(t)$

4 For  $i = 2, \dots, n + 1$ ,

$$\hat{I}_i(t) = I_i^*\left(t + \sum_{i=1}^{i-1} I_i(t)\right)$$

5 Return  $L(td)$  and  $\hat{I}_i(td) = d \times \hat{I}_i(t)$ ,  $i = 1, \dots, n + 1$

For  $i = 1, \dots, n + 1$ ,  $A_i = \lfloor a_i/d \rfloor$  and  $B_i = \lfloor b_i/d \rfloor$ , where  $[-a_i, b_i]$  is the range of allowed change to the  $i$ th IBI ( $I_i$ ). In each iteration of Step 2,  $F_i(td + H_i)$  should be computed and stored before the optimization.

spiking activity of the  $c$ th unit is a binary sequence  $x_c(1), x_c(2), x_c(3), \dots$ , such that

$$x_c(t) = \begin{cases} 1 & \text{if the unit generates spikes in the } t\text{th time bin,} \\ & \text{i.e. in the between } (t - 1)\Delta \text{ and } t\Delta \\ 0 & \text{otherwise.} \end{cases}$$

When  $\Delta$  is small enough, almost all time bins can have at most one spike, and the binary representation induces little loss of information on spike count.

To start, one needs to collect a sample of multielectrode spike trains associated with multiple occurrences of the event sequence. For each sample spike train, if  $T_1, \dots, T_n$  are the associated event times, then the segments of the spike train in surrounding intervals  $[T_i - A\Delta, T_i + B\Delta]$  are registered as codewords, where  $A, B \geq 0$  are fixed integers. As in the case of point process representation, the codewords cannot be warped in time.

STATISTICAL MODEL. The procedure is based on the following assumptions.

- Conditional on the occurrence of an event at time  $t$ , the spiking activities of the units in the time interval  $[t - A\Delta, t + B\Delta]$  are independent, such that the activity of each unit forms a Poisson process.
- Conditional on the occurrence of the entire event sequence, the joint spiking activities of the units associated with the individual events and the IEIs are all independent of each other.

The above assumptions of conditional independence are weaker than the assumption of unconditional independence because they require independence only during certain specific events. The assumption allows for continuity of the firing rates over time as well as different firing rate functions to be associated with different events.

For  $i = 1, \dots, n$ ,  $c = 1, \dots, C$ , and  $j = -A, \dots, B$ , let  $p_{i,c}(j)$  be the probability that the  $c$ th unit generates spikes in the  $j$ th time bin away from the  $i$ th event. The probability can be estimated using a peristimulus time histogram method. That is, for the  $i$ th event and  $c$ th unit, we first align the sample spike trains of the unit at the event, then estimate  $p_{i,c}(j)$  as the fraction of those sample spike trains that have spikes in the  $j$ th time bin relative to the time of the event.

For  $i = 1, \dots, n - 1$ , we use  $q(t, \theta_i)$  to model the probability density of the duration of the  $i$ th IEI. The parameter  $\theta_i$  can be estimated based on the sample as well.

LOCAL SCORES FOR SPIKING ACTIVITY PATTERNS AROUND EVENTS. To identify the occurrences of the whole event sequence, the first step is to evaluate the likelihood of the occurrence of a single event in each time bin. The derivation of the likelihood is given in the APPENDIX. Here we consider the computational procedure only. For the  $i$ th event and the  $c$ th unit, we introduce  $\Phi_{i,c}(j)$ , such that for  $j = -A, \dots, B$ , it is evaluated as

$$\Phi_{i,c}(j) = \log \frac{p_{i,c}(-j)}{1 - p_{i,c}(-j)}$$

It follows that

$$\log(\text{Prob}\{\text{the } i\text{th event occurs in the } t\text{th time bin}\}) = \sum_{c=1}^C \Phi_{i,c} * x_c(t) + \text{Const.} \quad (7)$$

where  $*$  denotes the mathematical convolution and  $\Phi_{i,c} * x_c(t)$  is the  $t$ th entry of the convolution between  $\Phi_{i,c}$  and  $x_c$ . The derivation of the formula is given in the APPENDIX. We therefore use

$$F_i(t) = \sum_{c=1}^C \Phi_{i,c} * x_c(t)$$

as the similarity score between the joint spiking activity starting in the  $t$ th time bin and the pattern associated with the  $i$ th event. The sequence  $\Phi_{i,c}$  acts as a discrete linear filter. By convolving  $\Phi_{i,c}$  with the entire binary sequence  $x_c$ , the score is obtained for all time bins.

GLOBAL SCORE FOR THE EVENT SEQUENCE. For the  $i$ th IBI,  $i = 1, \dots, n - 1$ , and  $t \geq 1$ , let  $G_i(t) = -\log q(t\Delta, \theta_i)$ . Then for any time bin  $t$  and  $I_1, \dots, I_{n-1}$ , the logarithm of the probability that the  $t$ th time bin contains the onset of an occurrence of the event sequence with IEIs  $I_1\Delta, \dots, I_{n-1}\Delta$  can be expressed as  $l(t, I_1, \dots, I_{n-1})$  plus a constant, where

$$l(t, I_1, \dots, I_{n-1}) = \sum_{i=1}^n F_i(t + I_1 + \dots + I_{i-1}) - \sum_{i=1}^{n-1} G_i(I_i) \quad (8)$$

We therefore define the score assigned to the time bin  $t$  as

$$L(t) = \max_{I_1, \dots, I_{n-1}} l(t, I_1, \dots, I_{n-1}) \quad (9)$$

Up to an additive constant,  $L(t)$  is the maximum likelihood that an occurrence of the event sequence starts in the  $t$ th time bin.

Denote by  $\hat{I}_1(t), \dots, \hat{I}_{n-1}(t)$  the maximizers associated with  $L(t)$ . As in the case of the point process representation,  $L(t), \hat{I}_1(t), \dots, \hat{I}_{n-1}(t)$  can be computed using DTW. The algorithm is sketched in Table 3.

FINDING OCCURRENCES OF EVENTS. Once the score function  $L(t)$  is computed for all time bins, we can use it to detect occurrences of the event sequence. The detection is based on the assumption that the score at the onset of an occurrence of the event sequence (i.e., the time of the first event) is higher than the scores elsewhere within the neighborhood of the onset. The local maximum points associated with  $L(t)$  are therefore identified as potential onsets of occurrences of the event sequence. For each identified  $t$ , the time bins of the onsets of the associated events are obtained as

$$t'_i = t + \hat{I}_1(t) + \dots + \hat{I}_{i-1}(t) \quad i = 1, \dots, n$$

The actual times of the events are estimated as  $t'_i\Delta$ . In practice, because the score function  $L(t)$  can be quite noisy, it may need to be smoothed to extract meaningful local maximum points.

TABLE 3. DTW for the binary representation

1 Construct filters  $\Phi_{i,c}$  for all  $i = 1, \dots, n$  and  $c = 1, \dots, C$

2 Compute

$$F_n(t) = \sum_{c=1}^C \Phi_{n,c} * x_c(t)$$

and define  $L_n(t) = F_n(t)$

3 For  $i = n - 1$  down to 1,

$$F_i(t) = \sum_{c=1}^C \Phi_{i,c} * x_c(t)$$

$$I_i^*(t) = \operatorname{argmax}_l [F_i(t) + L_{i+1}(t + l) - G_i(l)]$$

$$L_i(t) = F_i(t) + L_{i+1}(t + I_i^*(t)) - G_i(I_i^*(t))$$

4 Let  $L(t) = L_1(t)$  and  $\hat{I}_1(t) = I_1^*(t)$

5 For  $i = 2, \dots, n - 1$ ,

$$\hat{I}_i(t) = I_i^* \left( t + \sum_{k=1}^{i-1} \hat{I}_k(t) \right)$$

6 Return  $L(t)$  and  $\hat{I}_1(t), \dots, \hat{I}_{n-1}(t)$

To summarize, the numerical procedure for the binary representation starts with estimating  $p_{i,c}(j)$  and  $q(x, \theta_i)$ . Next the procedure constructs filters  $\Phi_{i,c}$  for  $i = 1, \dots, n$ ,  $c = 1, \dots, C$ , and also constructs  $G_i(x)$  for  $i = 1, \dots, n - 1$ . The scores and potential occurrences of the event sequence are then computed by the algorithm in Table 3.

## RESULTS

### Application to RA of zebra finch

We applied the procedure for point process representations to the SU spiking activity in RA of adult male zebra finch. The data were high signal-to-noise ratio recordings from chronically implanted movable electrodes in RA, collected in a previous study (Dave and Margoliash 2000). Male birds directed their songs toward females housed in adjacent half-cages.

For each RA unit, the goal was to identify segments in its spontaneous spiking activity during sleep that exhibited temporal patterns similar to its premotor activities associated with daytime singing. Such segments have been interpreted as replay of premotor activity during sleep that may be involved in learning and maintaining song (Margoliash 2002). The female-directed songs of adult male zebra finches are highly stereotyped intense behaviors (Sossinka and Böhner 1980), acquired through learning (Brenowitz et al. 1997), and are dynamically maintained in adults by auditory feedback (Nordeen and Nordeen 1992). The songs constitute sequences of "syllables" repeated in a fixed pattern to form one or more repeated "motifs." Premotor spiking activity in RA associated with song renditions are highly stereotyped and exhibit precise temporal patterning within bursts, with the SD in spike timing as low as 0.2 ms (Chi and Margoliash 2001; Yu and Margoliash 1996). Therefore the point process representation is preferred to the binary representation for RA activity.

For these data, the spontaneous activity of a unit during sleep can be probed with one or more templates. Once a

signature motif of the bird was identified, any spike train of the unit associated with a rendition of the motif could serve as a template (Fig. 2, A and B). To partition a template into bursts and IBIs, we aligned the spike trains associated with all the identified renditions of the motif by their onset times (Fig. 2C). The raster plot revealed common temporal structures of the spike trains that had a many-to-many correspondence to the acoustic structures of the motif. Based on the structures, we partitioned the spike trains into bursts and IBIs by the criteria that adjacent bursts be at least 20 ms apart. Within the bursts of these spike trains, the mean ISI was  $d' = 3.06 \pm 2.68$  ms, and within their IBIs the mean ISI was  $d = 73.02 \pm 55.97$  ms. The same partition was obtained when we applied a more refined alignment procedure which minimized the total  $L_1$  distances of the spike trains (Chi and Margoliash 2001).

We used the spike train in Fig. 2B as a template. The data of spontaneous activity during sleep had a duration of about 100 min, much longer than the duration of the template (660 ms). It would be very labor intensive to identify matching segments from the data by visual inspection. It is important to keep in mind that there were no behavioral events to serve as time referents for the sleep activity and so one could base the identification only on similarity scores.

To implement the procedure in Table 2, we used the bi-weight kernel to compute local similarity scores by Eq. 1. The parameter to control the temporal precision for spike matching was  $\Delta = d'/2 \times 1.875 = 3.06/2 \times 1.875 = 2.869$  ms by Eq. 5. To calculate the global similarity scores by Eq. 2, we set  $G_i(x) = 0$  and fixed the maximum proportion of warp at 1/5 for each template IBI. The score function was computed on a grid with step size 0.5 ms.

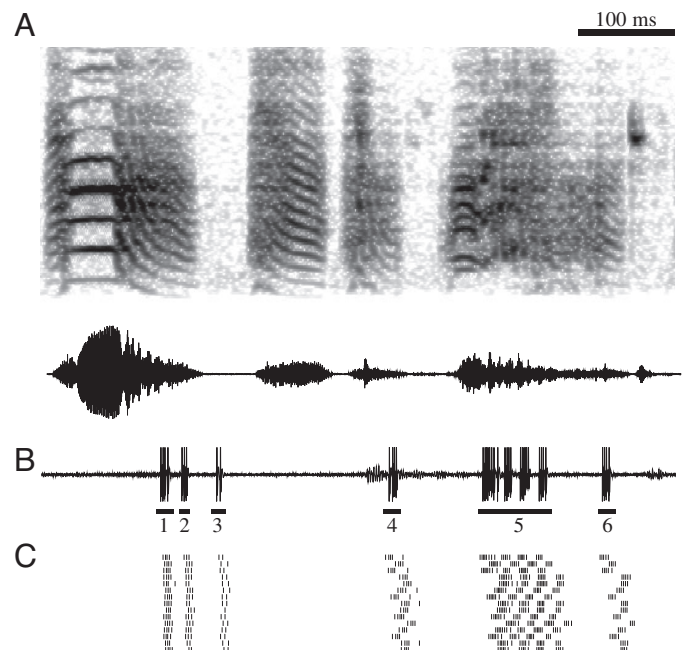


FIG. 2. A: spectrograph and waveform of a rendition of zebra finch song motif, which has a duration of 660 ms. B: a single-unit neuronal trace associated with the motif rendition. Trace has the same duration as the motif rendition, but starts 40 ms ahead. Lead time is taken out in the plot for the 2 to be aligned. C: raster plot of spike trains associated with different renditions of the same motif, aligned by the onset times of the renditions. Neuronal trace in B, in the 3rd row of the plot, is partitioned into 6 burst intervals marked by horizontal bars and 7 IBIs based on the raster plot.

During the spontaneous activity of the unit in sleep, the mean ISI was  $d_0 = 49.05 \pm 20.85$  ms. By Eq. 6, the penalty on each nonmatching spike was  $\nu = 0.1434$ . We then calculated the similarity scores across the data to identify segments that exhibited similar patterns to the template.

The results from one session of the data are shown in Fig. 3; plot A displays the similarity scores across. To identify matching segments, we set the threshold for similarity score  $\theta = N/3$ , where  $N = 41$  was the total number of spikes in the template. Four above-threshold local peaks were identified in this session, with scores 16.6, 19.2, 15.9, and 15.4, respectively. The second and fourth peaks are marked by  $x$  and  $z$  in Fig. 3A. By Eq. 4, we extracted the optimal matching segments at the peak locations and partitioned them into burst intervals and IBIs that were mapped one to one to those in the template (Fig. 3, B and C). The identified segments exhibited a high degree of similarity to the template and also showed nonuniform time warping of the IBIs.

The scores captured the similarity between the data segments and the template reasonably well. For example, in Fig. 3A, the score function peaks at location  $y$ , however, the value is only about half that at  $x$  and  $z$ . A closer look of the neuronal trace around  $y$  reveals that, although it has some similarity to the template, the level of similarity is substantially lower than those exhibited by the traces at  $x$  and  $z$ .

To see whether DTW was necessary in this example, we collected all the identified optimal matching segments from the data. There were 15 of them, with mean score  $16.6 \pm 1.9$ . To see how much warping of the template was involved to match

these segments, we analyzed the two longest IBIs in the template, which were the fourth and fifth IBIs. The time compressions/expansions of the two IBIs,  $-4.2 \pm 21.5$  and  $-2.5 \pm 9.6$  ms per segment, showed no significant correlation with each other ( $r = 0.34$ ,  $P = 0.22$ , 95% CI =  $[-0.21, 0.73]$ ,  $t$ -test). For instance, to match the segment at  $x$  in Fig. 3A, the two IBIs were compressed by 33.5 ms and expanded by 6.5 ms, respectively. This suggests that nonuniform, wide-ranging time warping was present in the data, supporting the use of DTW for the pattern identification.

Another question is whether the probability distribution of the time warping in the premotor activity could be used in the pattern identification for the sleep data. The answer seems to be negative. There is evidence that the variability of IBI in the spontaneous activity during sleep differed from that during wakeful behavior. In the preceding data, the variance of the fourth IBI in the identified matching segments was significantly larger than the one in the exemplar spike trains associated with motif renditions. The ratio of the two was  $r = 18.61$  ( $P < 10^{-5}$ , 95% CI =  $[6.25, 55.44]$ , F-test). The variances of the fifth IBI on the other hand did not exhibit significant difference ( $r = 2.20$ ,  $P = 0.15$ , 95% CI =  $[0.74, 6.54]$ ). The change in the statistical properties of neuronal activity across different behavioral states poses difficulty for the selection of parameter values, as we tried to use the templates obtained in one state to analyze the activity in another.

#### Sensitivity to temporal precision of spike matching

We had used  $\Delta = 1.875\delta$ , where  $\delta = 1.53$  ms was the half mean ISI within the bursts of the exemplar spike trains associated with the motif renditions, and the coefficient 1.875 “standardizes” the biweight kernel. To examine how sensitive the pattern identification was to  $\Delta$ , we compared the results to those obtained with different preselected values of  $\Delta = 1.875\delta$  while keeping the other parameters unchanged. For larger  $\delta$ , the temporal precision required for template matching was lower and more segments with above-threshold similarity scores were expected. The question is how sensitive the number of identified segments was to the change in  $\delta$ . Denote by  $n$  the number of segments that were identified using a preselected value of  $\delta$  but missed by the original procedure that used  $\delta = 1.53$  ms and  $n'$  the number of those the other way around

$\delta$ , ms	0.5	1	2	3	4
$n$	1	0	1	6	10
$n'$	15	3	0	0	0

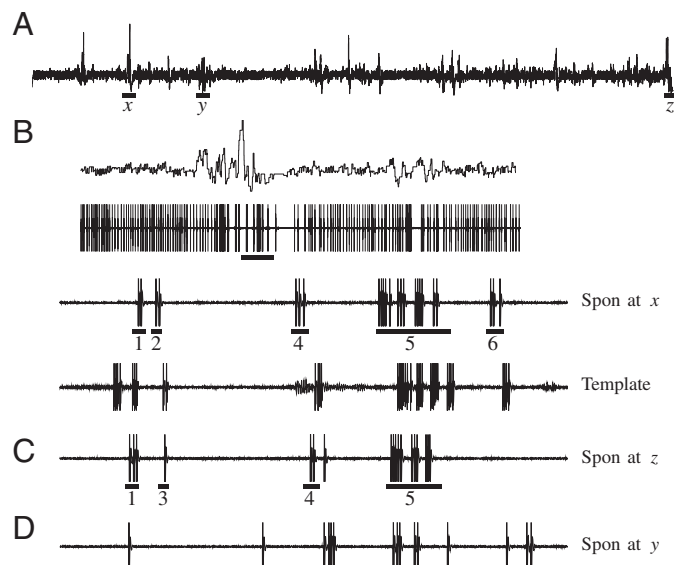


FIG. 3. Detection for RA activity. A: similarity scores over 5 min of the SU spontaneous activity in RA during sleep, using the spike train in Fig. 2B as the template. Three local peaks in the similarity scores are marked as  $x$ ,  $y$ , and  $z$ . B, top row: enlarged view of the scores in an interval of 11 s around  $x$ . 2nd row: RA activity in the same interval. Horizontal bar indicates a segment of RA activity that may have a pattern similar to that of the template. Onset of the segment is the location of the peak. 3rd row: an enlarged view of the segment of neuronal trace. Bursts in the segment are labeled with the indices of the corresponding optimal matching template bursts. A match to the 3rd template burst is missing. 4th row: same spike train in Fig. 2B shown for comparison. C: neuronal trace in the interval marked by  $z$ . Spike between the bursts labeled 4 and 5 is counted as a “noise” spike. D: neuronal trace in the interval marked by  $y$ . Relatively low score at  $y$  is indicative of the low level of similarity between the template and the activity in the interval.

Recall that the original procedure identified a total of 15 segments. Therefore the summary shows that the number of identified segments was reasonably stable for  $\delta$  around the value selected based on the exemplar spike trains. Except for  $\delta = 0.5$ , all the extra segments identified using the preselected values of  $\delta$  were not completely overlooked by the original procedure. They actually showed up as local peaks with relatively high scores ( $>10$ ). Therefore the increasing trend in  $n$  could be offset by lowering the threshold in the original procedure. The extra segment identified with  $\delta = 0.5$  was also not completely overlooked by the original procedure. It had roughly 50% overlap in time with a segment identified by the procedure.

### Sensitivity to the penalty on mismatching spikes

In the procedure, the penalty  $\nu = 0.1434$  was selected adaptively, based on the firing rates of the premotor activities associated with motif renditions and the spontaneous activity during sleep. We compared the results to those obtained with different fixed values of  $\nu$  while keeping the other parameters unchanged. Let  $n$  be the number of segments identified by using a preselected value of  $\nu$  but missed by using the adaptively selected value and  $n'$  the number of those the other way around:

$\nu$	0	0.25	0.5	0.75	1
$n$	2	0	0	0	0
$n'$	0	1	5	7	8

Recall that with the adaptively selected  $\nu = 0.1434$ , 15 optimal matching segments were identified. The summary shows that for values of  $\delta$  around the adaptively selected value, the identification results were quite similar.

### Sensitivity to kernel

We evaluated how sensitive the pattern identification procedure was to the kernel. We implemented the procedure using the square, triangular, and Epanechnikov kernels, respectively. By Eq. 5,  $\Delta$  had to be adjusted accordingly. Denote by  $n$  the number of segments that were identified using a different kernel but missed by using the biweight kernel and  $n'$  the number of those the other way around:

$K$	Square	Triangular	Epanechnikov
$\Delta$ , ms	1.53	3.06	2.30
$n$	4	2	0
$n'$	0	1	0

Recall that with the biweight kernel, 15 optimal matching segments were identified. Therefore the summary shows that the procedure was reasonably stable by using different kernels. The differences in the number of identified segments arose from the small changes in similarity scores. For example, all four extra segments identified using the square kernel actually generated local peaks in the score function obtained with the biweight kernel. The scores were just slightly below the threshold  $\theta = 41/3$ , with differences 0.54, 1.20, 1.47, and 0.50, respectively.

### Simulations for threshold selection and performance evaluation

For the pattern identification procedure, the issues of threshold selection and sensitivity to noise are complicated by the fact that the statistical properties of the sleep data are likely to be different in some unknown but important aspects from the premotor activity that generated the templates. The simulation method is based on the idea that by testing the pattern identification procedure on artificial data that incorporate certain statistical properties of the sleep data, some reasonable inference might obtain.

We generated simulated data using the exemplar spike trains associated with motif renditions (Fig. 2A) because they had unambiguous interpretation. The pattern identification procedure was then applied to the simulated data to yield score functions. Because the peak values in these score functions were associated with “true targets” (i.e., spike trains known to

exhibit the pattern of interest), we could use their distributions to draw inferences on the ability of the procedure to find true targets corrupted by noise.

Given an exemplar spike train  $T$  registered in  $[0, D]$ , a simulated spike train was generated as follows.

- Each spike in  $T$  was randomly deleted with probability  $q$  and each remaining spike was randomly shifted by  $s \sim N(0, \sigma)$ .
- The modified  $T$  was embedded into a larger interval  $I = [0, D + 2B]$  and a sample from the Poisson process with density  $\lambda$  was added to  $I$  as “noise” spikes.
- Finally, some of the spikes were removed from the resulting spike train so that all remaining ISIs were at least  $\tau$ .

Because DTW automatically accounts for time warping of IBIs, we did not randomly modify the IBIs. In all the simulations,  $\tau = 1$  ms, which is a plausible lower bound for the ISI in RA. We also fixed  $B = 500$  ms in all the simulations. The value of  $\lambda$  was selected based on the average firing rate of the spontaneous activity during sleep, which was estimated at 21 Hz. We set  $\lambda = 20$  Hz. The parameters  $q$  and  $\sigma$  are unobservable from the spontaneous activity during sleep. We set  $\sigma$  around or above the half-mean  $\delta (= 1.53$  ms) of the ISI within the bursts of the exemplar spike trains, so that there was a good chance for a spike in the exemplars to have a large “jitter” relatively to the ISIs. We conducted four simulations, with  $q = 1/4, 1/3$  and  $\sigma = 1.5, 2$  ms, respectively.

For each of the exemplar spike trains, 100 simulated spike trains were generated. The raster plot of a total of 1500 simulated spike trains with  $\sigma = 1.5$  ms and  $q = 1/3$  is shown in Fig. 4. Despite the relatively high chance of spike deletion in the true targets, the simulated spike trains maintained the overall temporal structure of the exemplar spike trains. However, not all of them exhibited enough similarity to the exemplars. A small percentage  $a$  of the simulated spike trains either failed to yield similarity scores above  $\theta = N/4$ , with  $N$  the number of spikes in the template, or failed to have peak locations in score function within 50 ms from the onsets of the true targets. Excluding these spike trains, denote by  $S$  the peak value in the score function of any remaining spike train. The distribution of  $S$  was close to a Normal distribution. As  $\sigma$  or  $q$  increased,  $a$  increased and the distribution of  $S$  shifted toward 0.

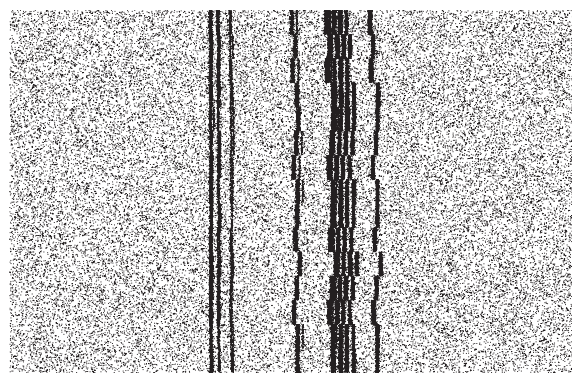


FIG. 4. Raster plot of simulated spike trains with “noise” added to randomly modified copies of the exemplar spike trains in Fig. 2A. Horizontal bar represents the time interval of 660 ms where the random copies are inserted. Each exemplar spike train generated 100 random spike trains that appear in the same block in the raster plot.

Simul.	$\sigma$ , ms	$q$	$a$	$\mu_S$	$\sigma_S$
1	1.5	1/4	$1.3 \times 10^{-2}$	18.20	2.51
2	1.5	1/3	$2 \times 10^{-2}$	17.17	2.44
3	2	1/4	$0.9 \times 10^{-2}$	17.58	2.53
4	2	1/3	$2.3 \times 10^{-2}$	16.48	2.40

A threshold could be selected based on the distribution of  $S$ . A sensible choice would be  $\mu_S - 2\sigma_S$ . For all the simulations, this value was a little less than the actually used threshold value  $N/3$ . Therefore a few more segments would have been identified, and the sum of  $a$  and the tail probability under  $N(\mu_S, \sigma_S)$  ( $\approx 2.5\%$ ) could serve as a nominal false negative rate.

We also tested with increased noise level  $\lambda = 50$  Hz. In these simulations, the signal-noise ratio became much lower, greatly reducing the number of identified segments. For example, with all the other parameters the same as in the second simulation, the number of identified segments was only 87 (results not shown). This result, however, made no sense for the data we analyzed. Our suggestion is that to conduct meaningful simulations, the parameters in the simulations should be tuned as much as possible according to the data.

#### Application to MI of macaque monkey

We applied the procedure for the binary representation of neuronal activity to recordings obtained from the arm area of MI in the left hemisphere of a trained adult macaque monkey. The data were collected using a chronically implanted silicon microelectrode array composed of 100 electrodes (1.0-mm electrode length; 400- $\mu\text{m}$  interelectrode separation; Cyberkinetics Neurotechnology System, Foxboro, MA). Based on histological evidence from previous implants that we have performed, the electrode tips are likely to be in lower portions of layer 3 or in layer 5. To acquire extracellular action potentials, signals were amplified (gain = 5,000), band-pass filtered (250 Hz to 7.5 kHz) and sampled at 30 kHz per channel. Only waveforms that crossed a threshold were stored and spike-sorted using Off-line Sorter (Plexon, Dallas, TX). The monkey was trained to repeatedly perform a sequence of movements (see following text) to receive a juice reward. From a decoding perspective, it is of interest to have a means to estimate the times of certain landmark events in the behavior during each trial based on the associated neuronal activity. These estimates were compared with the actual event times recorded during each trial.

The monkey was operantly trained to perform a target pursuit task by moving a cursor to a target with its right arm. The cursor and the target were displayed on a horizontal projection surface in front of the monkey. The monkey's arm rested on cushioned arm troughs secured to links of a two-joint robotic arm (KINARM system; Scott 1999) underneath the projection surface. The shoulder joint was abducted  $90^\circ$  such that shoulder and elbow flexion and extension movements were made in the horizontal plane. The target was programmed to appear on the four corners of a fixed square. A trial could start at any time after the end of the previous one. At the start of a trial, the target appeared at the upper right corner of the square and the monkey had to move the cursor to it. Once reached, the target jumped counterclockwise to the next corner and the monkey had to move the cursor again, and so on (Fig. 5). A trial was registered as successful only if the monkey reached the target at all of the four corners within 5 s. In the data we

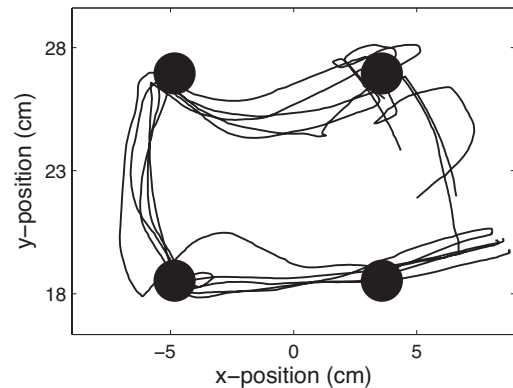


FIG. 5. Trajectories of hand movement of a macaque monkey in 5 successful trials of the target pursuit task as described under RESULTS. In each of the trials, the monkey's hand passed the 4 corners of a square in counterclockwise order to reach a target that jumped from one corner to another, starting from the top right corner. Time when the top right corner was reached was defined as the onset of the movement sequence in a trial. Corners of the square are denoted by circles.

collected, there were 240 successful trials; see Fig. 5 for some sample traces of hand movement in successful trials. A total of up to  $C = 49$  MI single units were simultaneously recorded.

In the experiment, a landmark event was defined as the cursor reaching any of the corners of the square which also corresponded to the time at which the next target appeared. A movement sequence started at the time when the first corner was reached. Therefore in a successful trial, there were  $n = 4$  events and three IEs. For each  $i = 1, 2, 3$ , the distribution of the duration of the  $i$ th IEI was concentrated between 0.5 to 1 s, but with a long tail extending  $\leq 3$  s. We fitted the distribution by the Gamma distribution

$$q(t, \theta_i) = \frac{1}{b_i^a \Gamma(a_i)} t^{a_i-1} e^{-t/b_i}$$

with parameter  $\theta_i = (a_i, b_i)$ . We shall mainly report results attained by incorporating the fitted Gamma distributions in the scoring. Later, we will show that similar results were attained as well without exploiting these distributions.

The temporal pattern of the spiking activity in MI has more variability than in RA. Taking this into account, the time bin duration was selected so that 99.9% of the time bins each contained at most one spike. Under this condition, the largest time bin duration was about 10 ms. We therefore chose  $\Delta = 10$  ms. The binning reduced the size of the data while keeping most of its information. The hand position data were also down-sampled at 10-ms resolution to estimate the times when the corners were reached.

We used the movement sequences in the first 200 successful trials as the training data set to estimate the firing probabilities  $p_{i,c}(j)$  within 1 s around each event, with  $i = 1, \dots, 4$  indexing the events,  $c = 1, \dots, 49$  indexing the units, and  $j = -100, \dots, 100$  indexing the time bins within 1 s around a time of event. The parameters  $\theta_i$  for the Gamma distributions were estimated as well. The estimates of  $p_{i,c}(j)$  and  $\theta_i$  were then used to construct the template. The recording after the 200th successful trials was used to test the accuracy of the identification procedure.

The score function  $L(t)$  was computed according to Table 3 over the entire recording and then smoothed with a low-pass filter with cutoff frequency 0.5 Hz. Because successful trials

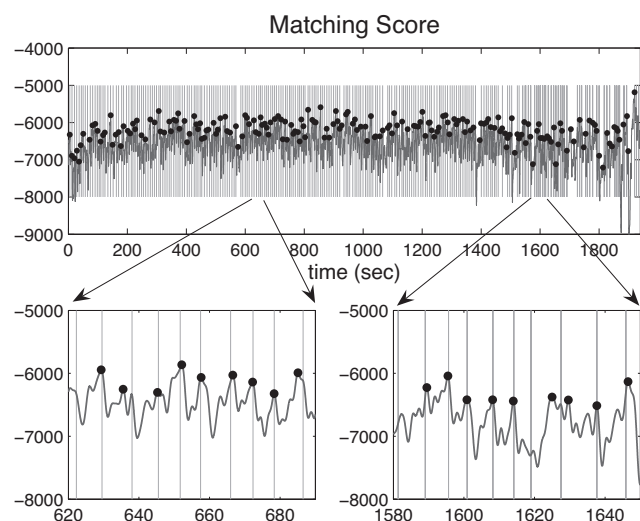


FIG. 6. Detection of movement sequences during successful trials in the target-pursuit task for the macaque monkey. *Top*: dark gray curve denotes the smooth score function using a low-pass filter on the original similarity scores over the entire recording period. Black dots mark all the estimated “max-maximum” points, which are local peak points of the smoothed score function with higher values than their adjacent local peaks. Gray vertical lines mark the actual onset times of movement sequences during successful trials (i.e., times of reaching the first corner). *Bottom*: sections extracted from the *top* panel showing details of the score function and identified onsets of movement sequences. *Left panel*: taken from the training data set (recording until the 200th successful trials). *Right panel*: test data set (the remaining 40 trials).

occurred at a much lower frequency ( $\leq 0.21$  Hz) than the cutoff frequency, the smoothing had no aliasing effect on the estimation of the onset times of the movement sequences (i.e., the times when the first corner was reached) in successful trials.

Across the entire recording period, 622 time locations  $t_i^*$  were identified where the smoothed score function  $\tilde{L}(t)$  peaked locally (Fig. 6). Based on the assumption that the scores were higher near the onsets of the movement sequences in successful trials than elsewhere, only  $t_i^*$  with  $\tilde{L}(t_i^*) > \tilde{L}(t_{i\pm 1}^*)$  were selected, resulting in 202 estimated onset times. For each of the selected  $t_i^*$ , the corresponding event times were estimated as in Table 3 (Fig. 7).

To evaluate the performance of the procedure, a selected  $t_i^*$  was defined as a true positive if the mean error of its associated estimated event times was  $< 1$  s; that is, the estimated event times were  $< 1$  s on average from the actual event times. Any selected  $t_i^*$  with the mean error  $\geq 1$  s was defined as a false positive. The counts of true positives, false positives, selected  $t_i^*$ , and events that actually occurred in the training data set, test data set, and the entire data set are as follows

	Training	Test	Total
True positives	144	29	173
False positives	18	11	29
Selected $t_i^*$	162	40	202
Successful trials	200	40	240

The power of the procedure was defined as the ratio of the number of true positives to that of successful trials. Evaluated over the entire data set, the power was high (173/240). The result was similar when the power was evaluated within the test data set only (29/40). The procedure also attained a high rate of true positives, defined as the ratio of the number of true

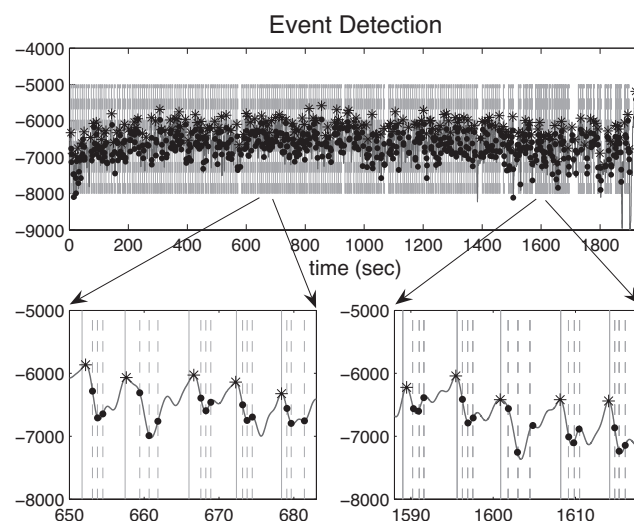


FIG. 7. Estimated times of events within identified movement sequences for the target-pursuit task experiment. In the *top* panel, the gray curve denotes the smoothed matching scores; the stars mark all the estimated starting points (first corner) of squared movements and the dots are the estimated points at the other 3 corners. Vertical lines mark the actual event times at the 4 corners of squared movements: the solid lines are for first corners and the dashed lines are for the other 3 corners. *Bottom* panels extracted from the *top* panel are 2 legible examples in the training (*left*) and test (*right*) data.

positives to the number of selected  $t_i^*$ . For the test data set, the rate of true positives was 29/40. (Incidentally, within the test data set, the number of selected  $t_i^*$  was 40, the same as the number of successful trials.) The rate of true positives for the training data set was even higher (144/162).

To see how sensitive the above evaluation of power was to the threshold for the mean error, we varied the threshold value from 0.5 to 1.5 s, causing the power to change from 62 to 73% (Fig. 8A). The slight change in the power indicates that the

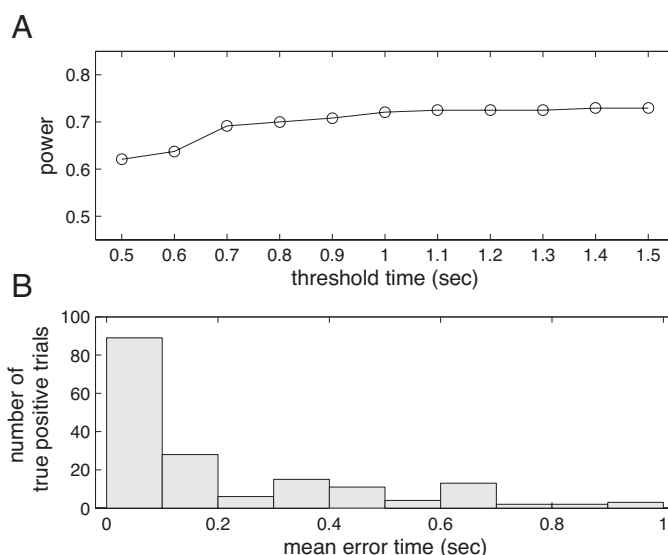


FIG. 8. *A*: robustness of power with respect to varied thresholds ranging from 0.5 to 1.5 s. Power varied from 62 to 73% in the examined range. *B*: histogram of mean error of estimated event times associated with the true positives identified from the entire recording. For each true positive, the mean error is the average of the errors in the estimates of the times when the 4 corners were reached. Among a total of 173 true positives, 144 were identified from the training data set and the remainder from the test data set.

evaluation was robust to the threshold selection thus justifies using 1 s as the threshold value.

On average, the accuracy of the estimated event times associated with true positives was high. For most of the true positives, the mean error of the estimated event times was  $\ll 1$  s (Fig. 8B). For the test data set, the mean error was  $0.223 \pm 0.200$  s. For the training data set, the mean error was  $0.204 \pm 0.242$  s. Interestingly, the estimated event times associated with different corners of the square exhibited different levels of accuracy. For the test data set, the errors associated with the four corners were  $0.505 \pm 0.606$ ,  $0.044 \pm 0.063$ ,  $0.089 \pm 0.164$ , and  $0.254 \pm 0.469$  s, respectively. For the training data set, the mean errors were  $0.564 \pm 0.740$ ,  $0.001 \pm 0.009$ ,  $0.001 \pm 0.013$ , and  $0.250 \pm 0.691$  s, respectively. Therefore the estimates of the times when the second or the third corner was reached were substantially more accurate and those for the first corner were the least accurate.

Finally, we tested the procedure without incorporating the estimated distributions of the IEs in scoring. Equivalently, all  $G_i$  were set to 0. For the test data set, we again attained a high power (29/40) and a high rate of true positives (29/41). The mean error in the estimated event times was  $0.226 \pm 0.196$  s and the errors associated with the four corners were  $0.536 \pm 0.607$ ,  $0.045 \pm 0.062$ ,  $0.092 \pm 0.189$ , and  $0.231 \pm 0.433$  s. The results for the entire data set and those for the training data set were similar to these reported results. The results were only marginally inferior to those attained by exploiting the estimated distributions of the IEs.

#### Comparisons with Kalman filter decoding method

A number of decoding methods have been proposed for MI of Macaque monkeys, including population vectors, linear filters, artificial neural networks, or Kalman filters (Carmona et al. 2003; Georgopoulos et al. 1986; Serruya et al. 2002; Wu et al. 2006). These methods have been demonstrated to be effective and accurate in estimating kinematic parameters such as hand movement direction or hand trajectory. However, none of them addresses the estimation of the timing of landmark events during a movement. This is largely attributed to the experiment paradigms these methods are designed for, which generate fairly straightforward hand movement of monkey, making it unnecessary to decode event timing during the movement. In contrast, in our squared movement paradigm, the monkey moved its hand following four corners of a square in each trial. In this paradigm, the times when the hand reached the corners provided important information on the movement. This raises a new perspective toward decoding: if a movement can be well defined by certain landmark events, then one may aim to decode the timing of the events without estimating the movement trajectory.

We compared the proposed procedure with the Kalman filter method, which was shown to yield better trajectory estimation than other current methods (Wu et al. 2006). We modified the Kalman filter procedure reported in Wu et al. (2006) so it could perform pattern identification. First, for each successful trial, we registered the time interval that started 1 s before the reaching the first corner and ended 1 s after the reaching the fourth corner. Then the procedure was applied to estimate the hand positions at discrete time points within the interval using the firing rates of all the 49 MI units. Finally, for each corner,

the nearest 5% estimated hand positions in terms of the Euclidean distance to the corner were identified, which accounted for about 25 time points per trial. The average of these time points was used as an estimate of the time when the corner was reached. In our study, the estimation was fairly robust when we varied the percentage of nearest hand positions from 5% to nearby values.

Note that unlike the template matching procedure, for the Kalman filter procedure, the spike trains had to be segmented into intervals containing individual trials. The segmentation was necessary to make sure that the Kalman filter procedure only used the nearest positions within the same trial to estimate the event times. On the other hand, the segmentation provided the Kalman filter procedure with critical information on time that was not accessible to the template matching procedure. Consequently, the estimated event times by the Kalman filter procedure were automatically within the trials, and so by our criteria on true positives, the procedure had power 1.

To make a reasonable comparison, we computed the mean errors in the estimated event times obtained by the Kalman filter procedure, in parallel to those obtained by the template matching procedure. Across the last 40 trials, which were automatically true positives for the Kalman filter procedure, the mean error was  $0.260 \pm 0.123$  s, and the errors associated with the four corners were  $0.262 \pm 0.178$ ,  $0.231 \pm 0.343$ ,  $0.184 \pm 0.156$ , and  $0.362 \pm 0.327$  s, respectively. Except for the estimated event times associated with the first corner, which benefited from the segmentation, the errors in the estimated event times associated with the other corners were significant greater than the template matching procedure. The same conclusion was drawn if we restricted to the 29 true positives identified by the template matching procedure. Across these trials, the Kalman filter procedure yielded mean error  $0.255 \pm 0.128$  s, and errors associated with the four corners  $0.269 \pm 0.166$ ,  $0.203 \pm 0.266$ ,  $0.184 \pm 0.164$ , and  $0.373 \pm 0.373$  s, respectively.

Essentially, the worse performance of the Kalman filter procedure was explained by the fact that it is intrinsically designed for the reconstruction of kinematic parameters, e.g., direction, speed, trajectories, and kinematic decoding and time decoding are rather different concepts. Figure 9 shows an example in which the Kalman filter procedure yielded a very good trajectory reconstruction but provided poor estimation for

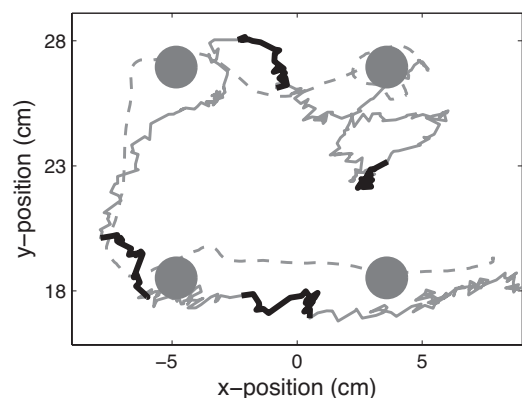


FIG. 9. One example trial: 4 corners of the square (gray circles), true hand trajectory (gray dashed line), and reconstruction using the Kalman filter (gray solid line). Thick black line denotes the reconstruction during a  $[-0.1$  s,  $0.1$  s] time window aligned with the true reaching time at each corner.

when the corners were reached. This example indicates that the estimation of timing has interest of its own and is worthy of investigation in the motor cortical study.

## DISCUSSION

### *Current pattern identification methods*

In neurophysiological analysis, various template matching methods have been developed to identify segments of patterned activity from neuronal recordings (Abeles et al. 1988, 1993; Chi et al. 2003; Dave and Margoliash 2000; Louie and Wilson 2001; Nádasdy et al. 1999, 2000). From the perspective that spike timing plays an important role in neural coding, most of the methods directly match spike times or ISIs registered on a continuum (Abeles et al. 1988, 1993; Dave and Margoliash 2000; Nádasdy et al. 1999, 2000). In these methods, similarity to a template is computed by counting the number of matching spikes or ISIs. Although the counting approach is suitable for patterns with a small number of spikes, in our experience, it significantly slows down for patterns with a moderate number of spikes, which is undesirable when relatively complex patterns need to be analyzed for large data sets. It is also not clear how to apply the approach effectively when simultaneous spiking activity of many units over hundreds of time bins needs to be compared with a template, as in the case of MI data.

When a pattern is represented by many spikes, template matching can be based on cross-correlating firing rate functions (Louie and Wilson 2001). Because this approach requires kernel smoothing, it is suitable for patterns without much fine temporal structure. However, as in the case of the RA data, when precise temporal structure is important for pattern identification, the firing rate–based approach is insufficient. The filtering method in Chi et al. (2003) is an attempt to balance between high temporal precision and computational efficiency. On the one hand, it evaluates similarity based on temporal closeness between data spikes and template spikes; on the other hand, it computes the similarity scores using a smooth profile of the template, in which each template spike is blurred into a “bump” that is a few milliseconds wide. Similarity scores across a data spike train are computed efficiently by linear convolution with the profile.

In all these methods, a template cannot be warped. To incorporate timescaling, the template has to be compressed or expanded uniformly in time beforehand. This is in contrast to the DTW approach of Victor and Purpura (1996), which adjusts template ISIs “on the fly” during template matching. This report proposes an approach that combines linear convolution and DTW for template matching. In this approach, linear convolution is applied at the level of individual components of a template, whereas DTW is applied at the level of the entire template. Because both linear convolution and DTW can be implemented efficiently, satisfactory computational efficiency can be achieved.

A combinatorial approach was recently proposed for pattern identification. The approach aims to identify data segments that exhibit a specific temporal order in which a set of neurons fire, regardless of the temporal details of the firing pattern (Lee and Wilson 2002, 2004; Smith and Smith 2006). This goal is different from the type of pattern identification considered here.

### *Flexible combination of local matching and time warping*

The proposed approach is flexible in modeling and combining information at two timescales. We showed that linear convolution can be applied for both point process representations and binary representations of spiking activity. Linear convolution can be applied for other representations as well. In general, as long as local similarity scores are defined as weighted sums of measures of spiking activity at different time points, they can be computed by linear convolution with a filter consisting of the weights involved. For example, to cross-correlate firing rates, the filter is simply the firing rate function of the template registered backward in time. In most cases, however, it is a major issue how to select appropriate filters. We showed that based on the principle of likelihood testing, some of the filter parameters can be estimated from data. The estimation-based approach reduces the number of parameters that have to be manually tuned, thus allowing a procedure to be more easily applied for different data sets and more objective comparisons of the results.

In our proposed pattern identification procedures, linear convolution and DTW are more or less two separable modules. Depending on the problem at hand, one may replace linear convolution with another method to compute local similarity scores, and then apply DTW to combine the local scores into similarity scores for the entire template. The point of using DTW is that it is capable of dealing with variability at a relatively large timescale. Such variability is evident in the spontaneous activity during sleep in RA (Dave and Margoliash 2000); it is even more evident in the analyzed MI activity, presumably because it was influenced by many factors. The distinction between local and more global temporal variabilities adapted herein is a simplification. Spiking activity can exhibit temporal variability at multiple timescales (Reich et al. 2000, 2001; Zugaro et al. 2005). From a computational perspective, to handle the variability at more finely graded timescales, one could introduce a hierarchy of procedures, each one computing the similarity scores at one timescale by combining scores obtained at a smaller timescale.

The DTW step in our approach has two useful properties. First, it estimates the times of individual components within each identified occurrence of a preselected pattern. The estimation of timing can be useful for decoding neural activity. Although our study does not preclude the possibility that the spike rates of neurons could be used to identify behaviorally relevant event times, there is a precedent for using spike patterns to identify the event times. Riehle et al. (1997) trained monkeys to expect go cues (i.e., cues to initiate movement) at certain predefined times. The study found that transient spike synchronization between two motor cortical neurons occurred at the expected cue times even when the cue did not appear. Interestingly, the spike rates of the individual neurons did not modulate during these expected times. One unanticipated result in our experiment was that the pattern identification procedure could attain very good estimates on when a macaque monkey moved a cursor to different targets. Further controlled experiments are necessary to determine the biological significance of this observation. For example, it is unclear whether these patterns related to aspects of the monkey's movement to the target or were associated with the visual appearance of the next target. Nevertheless, the example shows that, when appropri-

ately designed, a computational method can reveal intriguing phenomena that otherwise are difficult to detect. The discovery of very fine submillisecond neuronal “drift” in zebra finch RA spiking activity during singing is another such example (Chi and Margoliash 2001).

The second useful property of the DTW step is that it estimates the amount of warping for each compressible interval in the template. This may provide extra parameters to describe the large timescale dynamics of neural activity under certain condition. For example, as the experiment on the RA data indicates, one possibility is to test to what degree the assumption of uniform timescaling holds for the spontaneous spiking activity during sleep. On the other hand, although in principle extra parameters can provide more information for pattern identification, our experiment on the MI data indicates that it yielded only marginal improvement to incorporate a parametric statistical model for time warping. It remains to be seen whether parametric models can be more effectively exploited for pattern identification.

### *Selection of representation*

Although our pattern identification procedures for the point process representation and the binary representation appear to be quite different, they are based on the same perspective. That is, a spiking activity pattern can be divided into components that are incompressible in time and those that can be scaled in time, and the information about the pattern is mostly incorporated in the incompressible components. Most of the current pattern identification methods are based on point process representations. To avoid potential side effects arising from discretization of time, point process representations should be the default choice for pattern identification, even for spiking activity with higher (electrophysiological) variability than that in RA. Higher variability can be accommodated by lowering the temporal precision for spike matching or using multiple filters to account for different aspects of variability (Chi et al. 2003).

If the precision of spike time acquisition or computational load is of concern, then binary representations are more suitable. The choice between the two types of representations should be problem dependent. For the MI data, because the goal was to estimate behavioral timing of a monkey, which intrinsically could be highly variable, it was reasonable to discretize spike times to the extent that the resulting loss of temporal information caused no abrupt changes in the estimation. The binary representation was tested with multiple SUs extracted from simultaneous records drawn from an electrode array. It should be noted that continuous multiunit (MU) data, as opposed to single-unit or multielectrode data, can also be transformed into binary representations. Depending on the quality of the MU recording, it may be advantageous to set an amplitude threshold, count the number of threshold crossings, and quantify this as a binary signal. Alternatively, for recordings that sum over a larger number of units, it may be advantageous to measure the total amplitude of the rectified signal over some bin (e.g., Margoliash 1986) and quantify this as a binary signal. The pattern identification approach may then be applied to the binary signal.

### *Limitations and possible improvements*

One important situation that the proposed approach cannot handle is where the relative timing of the spiking activity of multiple units exhibit variability. Such “interunit temporal variability” may be significant if the recorded units are in different nuclei. Our procedure for the point process representation of neuronal activity is applicable only for SU recordings. The procedure for the binary representation essentially assumes that the units are time-locked with each other in their spiking. Therefore it cannot apply when the interunit timing variability has to be taken into account. One possible approach to this type of variability is to treat the relative timing of the units the same way as IBIs with appropriate DTW procedures. However, even with a moderate number of recorded units, the computational load can become heavy because the optimization involved in DTW may have to take into account the relative timing for many pairs of units. An intermediate approach may be to group the units into clusters, such that within each cluster, the interunit variability can be ignored. Then, appropriate DTW procedures may be developed for the clusters.

A second issue is how to combine the two pattern identification procedures developed here. This can be useful when different classes of neurons exhibit different levels of temporal precision in their spiking activity. For example, in the birdsong system, the spiking activity in the nucleus HVC, which is afferent to RA, has projection neurons with high temporal precision and interneurons with lower temporal precision (Hahnloser et al. 2002; Yu and Margoliash 1996). In general, given the interest in simultaneous recordings from multiple areas (Hoffman and McNaughton 2002), it is valuable to develop pattern identification procedures for the joint activity in multiple areas. To manipulate behavior directly based on spiking activity, it would also be useful to extend our procedures for real-time pattern identification.

The proposed pattern identification approach does not offer a recipe for threshold selection for similarity scores. The simulation-based method in the experiment for RA could provide a reasonable range for the threshold values. However, because no simulations can account for all the statistical properties of spiking activity, simulation methods should be used with caution. Despite its statistical derivation, the proposed approach is essentially of observational nature. In the end, how large a similarity score should be to identify a data segment as a match is based on an experimenter’s own judgment, irrespective of whether it is supported by statistical significance measures such as *P* values. The biological significance of the identified patterns can be determined only by controlled experiments.

## APPENDIX

### *Score function for the point process representation*

The score function is derived from a likelihood ratio test on the pattern of interest versus background. We model the background as a Poisson process with constant density  $\lambda_0$ . Let  $T = \{t_1, t_2, \dots\}$  be a template representing the pattern of interest. It randomly generates a data spike train as follows.

- 1) Each  $t$  in  $T$  generates a set of spikes from a constrained Poisson process with density  $g(x - t)$ , such that the likelihood of  $t$  generating  $A = \{x_1, \dots, x_k\}$  is

$$p(A) = \begin{cases} Z^{-1} \prod_{i=1}^k g(x_i - t) & \text{if } 0 \leq k \leq N \text{ and } |x_i - t| < \Delta \\ 0 & \text{otherwise} \end{cases}$$

Here  $\delta$  is the maximum “jitter” in spike times,  $N$  is the maximum number of spikes a template spike can generate, and  $Z = Z_N$  is a normalizing constant. Although  $N$  depends on the recorded unit, it will be seen that different values of  $N$  lead to the same form of similarity scores up to an additive constant. Therefore we can use the same formula for template matching without having to estimate  $N$ .

2) The IBIs are compressed or expended independently of each other, such that the amount of change to the  $i$ th IBI follows a probability distribution  $f_i(x)$ .

3) For each warped IBI, a sample from the Poisson process with density  $\lambda < \lambda_0$  is added into the warped IBI as noise spikes. The density of spikes within the pattern is assumed to be continuous. Therefore  $\lambda = g(-\Delta) = g(\Delta)$ .

For a spike train  $S$  that starts at time  $x$  and consists of spikes  $s_1 < s_2 < \dots$ , the log-likelihood that it is generated from  $T$  in such a way that the IBIs of  $T$  are changed by  $v_1, \dots, v_{n+1}$  and the  $(h_{i-1} + 1)$ th,  $\dots$ ,  $(h_i)$ th spikes in  $S$  are generated by  $t_i$  is

$$L_1 = \sum_{i=1}^{n+1} \log f_i(v_i) + \log \lambda \times (\text{total number of spikes in the warped IBs of } T) \\ + \sum_i \left[ \log \frac{1}{Z} + \sum_{j=h_{i-1}+1}^{h_i} \log g(s_j - x - t_i - v_1 - \dots - v_{k_i}) \right]$$

where  $k_i$  is the index of the template burst that contains  $t_i$ . On the other hand, the log-likelihood of  $S$  being part of the background is  $L_0 = \log \lambda_0 \times (\text{total number of spikes in } S)$ . The log-likelihood ratio is then

$$\sum_{i=1}^{n+1} \log f_i(v_i) + \log \frac{\lambda}{\lambda_0} \times (\text{total number of spikes in the warped IBs of } T) \\ + \sum_i \sum_{j=h_{i-1}+1}^{h_i} \log \frac{g(s_j - x - t_i - v_1 - \dots - v_{k_i})}{\lambda_0} + C$$

with  $C$  a constant dependent only on  $T$ . Next write  $\log [g(x)/\lambda_0]$  as  $(1 + \nu)K(x/\Delta) - \nu$ , where  $\nu \geq 0$  and  $0 \leq K(x) \leq 1$  is defined on  $[-1, 1]$  with  $K(0) = 1$ ,  $K(\pm 1) = 0$ . To this end, divide the log-likelihood ratio by  $M = \log [g(0)/\lambda_0]$ , and define

$$\nu = \frac{\log (\lambda_0/\lambda)}{M} = \frac{\log \lambda_0 - \log \lambda}{\log g(0) - \log \lambda_0}, K(x) = \frac{\nu + \log g(\Delta x)/M}{1 + \nu} \quad (\text{A1})$$

Let  $G_i(\nu) = -\log f_i(\nu)/M$  and ignore constant  $C$ . Then, following Chi et al. (2003), the transformed log-likelihood ratio is  $l(x, v_1, \dots, v_{n+1})$  in Eq. 2. As a result, up to a multiplicative factor  $M$  and an additive factor  $C$ ,  $L(x)$  in Eq. 3 is the maximum log-likelihood ratio of  $S$  being generated from  $T$  versus from the background.

### Score function for the binary representation

We first derive the score function associated with a single event, i.e., Eq. 7. For each  $t = 1, 2, \dots$ , let

$$P_i(t) = \text{Prob}\{\text{the } i\text{th event occurs in the } t\text{th time bin}\}$$

Then under the assumptions of conditional independence in our statistical model

$$P_i(t) = \prod_{c=1}^C \prod_{j=-A}^B P_{i,c}(j)^{x_c(t+j)} [1 - p_{i,c}(j)]^{1-x_c(t+j)}$$

and thus

$$\begin{aligned} \log P_i(t) &= \sum_{c=1}^C \sum_{j=-A}^B \{x_c(t+j) \log p_{i,c}(j) + [1 - x_c(t+j)] \log (1 - p_{i,c}(j))\} \\ &= \sum_{c=1}^C \sum_{j=-A}^B x_c(t+j) \log \frac{p_{i,c}(j)}{1 - p_{i,c}(j)} + \sum_{c=1}^C \sum_{j=-A}^B \log (1 - p_{i,c}(j)) \\ &= \sum_{c=1}^C \sum_{j=-A}^B x_c(t+j) \Phi_{i,c}(-j) + C_i \\ &= \sum_{c=1}^C \Phi_{i,c} * x_c(t) + C_i \end{aligned}$$

where  $C_i$  is a constant dependent only on  $i$ . This therefore verifies Eq. 7.

To verify Eq. 8, following the same idea for the point process representation (cf. Fig. 1B), it can be seen that the probability that the  $t$ th time bin contains the onset of an occurrence of the event sequence with IEIs  $I_1\Delta, \dots, I_{n-1}\Delta$  is

$$\begin{aligned} P(t, I_1\Delta, \dots, I_{n-1}\Delta) \\ &= \prod_{i=1}^n \text{Prob}\{\text{the } i\text{th event occurs in the } (t + I_1 + \dots + I_{i-1})\text{th time bin}\} \\ &\quad \times \prod_{i=1}^{n-1} q(I_i\Delta, \theta_i) = \prod_{i=1}^n P_i(t + I_1 + \dots + I_{i-1}) \times \prod_{i=1}^{n-1} q(I_i\Delta, \theta_i) \end{aligned}$$

Take the logarithm on both sides. Then

$$\log P(t, I_1\Delta, \dots, I_{n-1}\Delta) = \sum_{i=1}^n F_i(t + I_1 + \dots + I_{i-1}) - \sum_{i=1}^n G_i(I_i) + C$$

with  $C$  a constant. This therefore verifies Eq. 8. The interpretation of  $L(t)$  in Eq. 9 as a maximum likelihood up to an additive constant follows from a similar argument as the one for the point representation. For brevity, the argument is not elaborated.

### ACKNOWLEDGMENTS

We thank D. Paulsen for training the monkeys and collecting the data.

### GRANTS

This work was partially supported by National Institutes of Health Grants MH-68028 and DC-007206 to D. Margoliash and Z. Chi, and a grant from the Whitehall Foundation and National Institute of Neurological Disorders and Stroke Grants N01-NS-2-2345 and NS-45853 to N. G. Hatsopoulos.

### REFERENCES

- Aach J, Church GM. Aligning gene expression time series with time warping algorithms. *Bioinformatics* 17: 495–508, 2001.
- Abeles M, Bergman H, Margalit E, Vaadia E. Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *J Neurophysiol* 70: 1629–1638, 1993.
- Abeles M, Gerstein GM. Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. *J Neurophysiol* 60: 909–924, 1988.
- Anderson SE, Dave AS, Margoliash D. Template-based automatic recognition of birdsong syllables from continuous recordings. *J Acoust Soc Am* 100: 1209–1219, 1996.
- Bishop MJ, Thompson E. Maximum likelihood alignment of DNA sequences. *J Mol Biol* 190: 159–165, 1986.
- Brenowitz EA, Margoliash D, Nordeen KW. An introduction to birdsong and the avian song system. *J Neurobiol* 33: 495–500, 1997.
- Brockwell AE, Rojas AL, Kass RE. Recursive Bayesian decoding of motor cortical signals by particle filtering. *J Neurophysiol* 91: 1899–1907, 2004.
- Buzsáki G. Two-stage model of memory trace formation: a role for “noisy” brain states. *Neuroscience* 31: 551–570, 1989.

- Buzsáki G.** Memory consolidation during sleep: a neurophysiological perspective. *J Sleep Res* 7, Suppl. 1: 17–23, 1998.
- Carmena JM, Lebedev MA, Crist RE, O'Doherty JE, Santucci DM, Dimitrov DF, Patil PG, Henriquez CS, Nicolelis MAL.** Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol* 1: 193–208, 2003.
- Chi Z, Margoliash D.** Temporal precision and temporal drift in brain and behavior of zebra finch song. *Neuron* 32: 899–910, 2001.
- Chi Z, Rauske PL, Margoliash D.** Pattern filtering for detection of neural activity, with examples from HVC activity during sleep in zebra finches. *Neural Comput* 15: 2307–2338, 2003.
- Dave AS, Margoliash D.** Song replay during sleep and computational rules for sensorimotor vocal learning. *Science* 290: 812–816, 2000.
- Georgopoulos AP, Schwartz AB, Kettner RE.** Neuronal population coding for movement direction. *Science* 233: 1416–1419, 1986.
- Hahnloser RHR, Kozhevnikov AA, Fee MS.** An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419: 65–70, 2002.
- Hahnloser RHR, Kozhevnikov AA, Fee MS.** Sleep-related neural activity in a premotor and a basal-ganglia pathway of the songbird. *J Neurophysiol* 96: 794–812, 2006.
- Hatsopoulos NG, Joshi L, O'Leary JG.** Decoding continuous and discrete motor behavior from motor and premotor cortical ensembles. *J Neurophysiol* 92: 1165–1174, 2004.
- Hoffman KL, McNaughton BL.** Coordinated reactivation of distributed memory traces in primate neocortex. *Science* 297: 2070–2073, 2002.
- Huang S-S, Gray RM.** Spellmode recognition based on vector quantization. *Speech Recogn* 7: 41–53, 1988.
- Ito K, Mori K.** Dynamic programming matching as a simulation of budgerigar contact-call discrimination. *J Acoust Soc Am* 105: 552–559, 1999.
- Kogan JA, Margoliash D.** Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: a comparison study. *J Acoust Soc Am* 103: 2185–2196, 1998.
- Lee AK, Wilson MA.** Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* 36: 1183–1194, 2002.
- Lee AK, Wilson MA.** A combinatorial method for analyzing sequential firing patterns involving an arbitrary number of neurons based on relative time order. *J Neurophysiol* 92: 2555–2573, 2004.
- Louie K, Wilson MA.** Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron* 29: 145–156, 2001.
- Margoliash D.** Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *J Neurosci* 6: 1643–1661, 1986.
- Margoliash D.** Evaluating theories of bird song learning: implications for future directions. *J Comp Physiol A Sens Neural Behav Physiol* 188: 851–866, 2002.
- Margoliash D, Staicer CA, Inoue SA.** Stereotyped and plastic song in adult indigo buntings, passerina cyanea. *Anim Behav* 42: 367–388, 1991.
- Nádasdy Z.** Spike sequences and their consequences. *J Physiol (Paris)* 94: 505–524, 2000.
- Nádasdy Z, Hirase H, Czurkó A, Csicsvari J, Buzsáki G.** Replay and time compression of recurring spike sequences in the hippocampus. *J Neurosci* 19: 9497–9507, 1999.
- Needleman S, Wunsch C.** A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 48: 443–453, 1970.
- Nordeen KW, Nordeen EJ.** Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches. *Behav Neural Biol* 57: 58–66, 1992.
- O'Shaughnessy D.** *Speech Communication: Human and Machine* (2nd ed.). New York: Wiley-IEEE Press, 1999.
- Rabiner L, Juang BH.** *Fundamentals of Speech Recognition*. Lebanon, IN: Prentice Hall, 1993.
- Reich DS, Mechler F, Purpura KP, Victor JD.** Interspike intervals, receptive fields, and information encoding in primary visual cortex. *J Neurosci* 20: 1964–1974, 2000.
- Reich DS, Mechler F, Victor JD.** Independent and redundant information in nearby cortical neurons. *Science* 294: 2566–2568, 2001.
- Riehle A, Grün S, Diesmann M, Aertsen A.** Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* 278: 1950–1953, 1997.
- Sankoff D, Kruskal JB.** *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*. Reading, MA: Addison-Wesley, 1983.
- Scott SH.** Apparatus for measuring and perturbing shoulder and elbow joint positions and torques during reaching. *J Neurosci Methods* 89: 119–127, 1999.
- Serruya MD, Hatsopoulos NG, Paninski L, Fellows MR, Donoghue JP.** Brain-machine interface: instant neural control of a movement signal. *Nature* 416: 141–142, 2002.
- Skaggs WE, McNaughton BL.** Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science* 271: 1870–1873, 1996.
- Smith AC, Brown EN.** Estimating a state-space model from point process observations. *Neural Comput* 15: 965–991, 2003.
- Smith AC, Smith P.** A set probability technique for detecting relative time order across multiple neurons. *Neural Comput* 18: 1197–1214, 2006.
- Sossinka R, Böhner J.** Song types in the zebra finch (*poephila guttata castanotis*). *Z Tierpsychol* 53: 123–132, 1980.
- Victor JD, Purpura KP.** Nature and precision of temporal coding in visual cortex: a metric-space analysis. *J Neurophysiol* 76: 1310–1326, 1996.
- Wagner RA, Fischer MJ.** The string-to-string correction problem. *J Assoc Comput Mach* 21: 168–173, 1974.
- Waterman MS, Gordon L, Arratia R.** Phase transitions in sequence matches and nucleic acid structure. *Proc Natl Acad Sci USA* 84: 1239–1243, 1987.
- Wilson MA, McNaughton BL.** Reactivation of hippocampal ensemble memories during sleep. *Science* 265: 676–679, 1994.
- Wu W, Black MJ, Mumford D, Gao Y, Bienenstock E, Donoghue JP.** Modeling and decoding motor cortical activity using a switching Kalman filter. *IEEE Trans Biomed Eng* 51: 933–942, 2004.
- Wu W, Gao Y, Bienenstock E, Donoghue JP, Black MJ.** Bayesian population decoding of motor cortical activity using a Kalman filter. *Neural Comput* 18: 80–118, 2006.
- Yu AC, Margoliash D.** Temporal hierarchical control of singing in birds. *Science* 273: 1871–1875, 1996.
- Zugaro MB, Monconduit L, Buzsáki G.** Spike phase precession persists after transient intrahippocampal perturbation. *Nat Neurosci* 8: 67–71, 2005.