Pattern Filtering for Detection of Neural Activity, with Examples from HVc Activity During Sleep in Zebra Finches

Zhiyi Chi
chi@galton.uchicago.edu
Department of Statistics, Committee on Computational Neuroscience,
University of Chicago, Chicago, IL 60637, U.S.A.

Peter L. Rauske
pete@drozd.uchicago.edu
Daniel Margoliash
dan@bigbird.uchicago.edu
Department of Organismal Biology and Anatomy, Committee on Computational
Neuroscience, University of Chicago, Chicago, IL 60637, U.S.A.

The detection of patterned spiking activity is important in the study of neural coding. A pattern filtering approach is developed for pattern detection under the framework of point processes, which offers flexibility in combining temporal details and firing rates. The detection combines multiple steps of filtering in a coarse-to-fine manner. Under some conditional Poisson assumptions on the spiking activity, each filtering step is equivalent to classifying by likelihood ratios all the data segments as targets or as background sequences. Unlike previous studies, where global surrogate data were used to evaluate the statistical significance of the detected patterns, a localized p-test procedure is developed, which better accounts for firing modulation and nonstationarity in spiking activity. Common temporal structures of patterned activity are learned using an entropy-based alignment procedure, without relying on metrics or pair-wise alignment. Applications of pattern filtering to single, presumptive interneurons recorded in the nucleus HVc of zebra finch are illustrated. These demonstrate a match between the auditory-evoked response to playback of the individual bird's own song and spontaneous activity during sleep. Small temporal compression or expansion, or both, is required for optimal matching of spontaneous patterns to stimulus-evoked activity.

1 Introduction

In analysis of neural activity, an important problem is detection of spike sequences with certain temporal or spatiotemporal patterns. In some situations, the pattern of interest is prespecified, as exhibited, for example, by the activity of a neuron during sensory stimulation or motor behavior. In general, the goal here is to recognize spiking activity of the same neuron at a different time or in a different state that exhibits patterns similar to the prespecified one. This case is becoming increasingly important in the study of sleep, where recent results indicate that spiking activity during reinforced behavioral learning or behavior that requires sensory feedback was "replayed" in spontaneous activity during sleep (Nádasdy, Hirase, Czurkó, Csicsvari, & Buzsáki, 1999; Dave & Margoliash, 2000; Louie & Wilson, 2001). Results of this sort provide evidence for the role of sleep in learning and memory consolidation that is complementary to studies of how sleep and behavior modify the correlations between neurons (Wilson & McNaughton, 1994; Skaggs & McNaughton, 1996). This letter focuses principally on detection with prespecified patterns.

In other situations, no pattern is prespecified. Instead, the goal is to recognize any activity pattern that occurs at a frequency above chance level. Detection of such patterns has been important in the study of the nature of the neural code, especially in assessing temporal and rate coding (Abeles & Gerstein, 1988; Abeles, Bergman, Margalit, & Vaadia, 1993; Riehle, Grün, Diesmann, & Aertsen, 1997; Date, Bienenstock, & Geman, 1998; Baker & Lemon, 2000). Detection of "excessive" spike patterns should be a stronger method than are correlational methods (Vaadia et al., 1995). In principle, all possible spike sequences have to be accounted for. In practice, this can result in a serious computational problem. To reduce the complexity of the problem, therefore, typically only sequences with a fairly small number of spikes have been considered. This computational constraint can have serious consequences in cases where coding of natural behavior is under consideration, since target sequences can extend for many seconds of behavior and dozens, if not hundreds, of spikes. The techniques derived in this letter are computationally efficient and may therefore also be of value for this second case. Here we do not directly address this problem, however.

To detect patterned spike sequences efficiently and reliably, template matching algorithms have been developed. In these algorithms, exemplar spike sequences observed during specific behaviors are used as templates. A commonly used algorithm is the "sliding sweeps" algorithm (Nádasdy et al., 1999; Abeles & Gerstein, 1988; Abeles et al., 1993; Dave & Margoliash, 2000). In this algorithm, for each iteration, the segment within a time frame is compared with the template by counting the number of matching spikes or interspike intervals (ISIs). This is slow when the template has many spikes, resulting in inefficient detection for large data sets. Because the detection treats matching spikes or ISIs equally, this technique is also insensitive to the temporal variability of spikes. To expedite detection, one may instead compare the firing-rate functions of the data and the template by cross-correlation (Louie & Wilson, 2001). This method requires several steps of normalization and kernel smoothing, which makes it also insensitive to the temporal details of spike sequences.

In this letter, we approach the detection of spiking activity as a special case of pattern detection for point processes. This framework provides an effective way to incorporate temporal detail and firing rates of spiking activity into pattern detection. It formulates the detection process as classification based on the likelihood ratio test. Under certain assumptions of the point processes, the classification is equivalent to linear convolution of the entire data set with a filter determined by the template. The filtering leads to substantial computational efficiency, as it allows detected patterns ("targets") to be located by above-threshold peaks in the filter response. The quality of the detection critically depends on the variability of the spiking activity. This is addressed using multiple filters and templates and at multiple timescales. The statistical significance of the targets is evaluated by tests using short processes to model the spiking activity around the targets. Such localized tests address different issues than the more commonly applied tests using long processes to model the entire data set (Abeles & Gerstein, 1988; Abeles et al., 1993; Riehle et al., 1997; Date et al., 1998; Baker & Lemon, 2000). Finally, in our approach, the detected targets and the templates can be compared in their common temporal structure, after learning the temporal structure of the templates.

We applied this approach to a small data set of single neurons recorded in the song control nucleus HVc of sleeping zebra finches. Sensorimotor matching had previously been reported for neurons in the nucleus robustus archistriatalis (RA), to which HVc projects (Dave & Margoliash, 2000). Here, we demonstrate a match between each HVc neuron's auditory-evoked response to playback of the individual bird's own song and the spontaneous activity of that neuron. Investigating such matching is important for theories that posit a role of sleep in birdsong learning (Margoliash, 2002).

Following is an overview of the main points of the filtering approach:

1. **Conditional Poisson assumption**: The filtering approach assumes conditional independence for the point processes. In terms of neural activity, this means that each finite spike train is randomly generated by a template or by the background, and for either case, the distribution is Poisson. By dividing spiking activity into different categories and approximating each with a Poisson process, the resulting "hybrid" process, which in general is not Poisson, can provide a better model of the spiking activity than Poisson processes. The approach is related to the generalized Hough transform (Rojer & Schwartz, 1992; Amit & Geman, 1999), with the key difference that it takes into account various models for the target sequences as well as the background.

2. **Classification and filtering**: A spike sequence is classified as a target instead of a background sequence only if the ratio of the corresponding likelihoods is above threshold in all tests. By the above model, the likelihood ratio is determined by a linear convolution of the sequence with a filter. The filter not only limits the amount of temporal discrepancy, or "jitter,"

between matching spikes, but also weighs spikes differentially, favoring spikes with small jitter. Henceforth, this convolution will be referred to as *pattern filtering*.

3. **Multiple filters**: Since each filter is based on a reasonable but fairly simple probability model of the spiking activity, it often detects targets that are not very similar to the template. This limitation is addressed by multiple filters, which implicitly induce a probability model subject to more constraints and hence offer a better description of the spiking activity. Computational efficiency is maintained by coarse-to-fine detection (Fleuret & Geman, 2001).

This approach is in contrast to using more sophisticated models. More sophisticated models reduce computational efficiency, especially when complicated patterns must be detected from large data sets. Also, more sophisticated models do not necessarily result in noticeable improvement in detection performance, yet they suffer from potential difficulty of parameter estimation.

4. **Multiple templates**: A challenge for spike pattern detection is variability in target sequences. Spike trains, even when associated with the same stimulus or behavior, often exhibit variability. To allow for this, multiple templates are used. Detection is first conducted using individual templates, independently of the other templates. Then the detected targets across the templates are combined.

5. **Multiple scales**: Systematic changes in spiking activity can also present problems. In some experimental designs, templates collected in one behavioral state are used to analyze the activity in a different state. There can be systematic differences between states in the rate that processes evolve; these differences are not accounted for by the templates (Wilson & McNaughton, 1994; Nádasdy et al., 1999; Dave & Margoliash, 2000; Louie & Wilson, 2001). To address this, filters are scaled in time by varying amounts, so that sequences with similar patterns as the templates up to a temporal scale can be detected.

6. **Training**: Some parameters in the detection process are selected by training. In some cases, such as the spontaneous activity during sleep, one has to account for the absence of reliably tagged spike sequences to be used as training sets. The solution here is to train the detector using simulated spike sequences based on biologically plausible parameters and then evaluate the sensitivity of the detection to the parameters.

7. **Significance test using simulation**: Assessing the p-values of targets poses a challenging problem, as it is difficult to model the probability distribution of the data. We note that the statistical significance of each target depends on only the properties of the data around it rather than the global properties of the entire data set. Therefore, we develop “local” tests, which use simulated spike trains based on the activity around the target. Using local processes, it is easier to account for the rate modulation and nonstationarity of the data.

The local tests, however, do not address questions related to the entirety of the data. Questions such as the statistical significance of the total number of detected targets require global methods, such as random shuffling tests (Abeles & Gerstein, 1988) and unitary event analysis (Grün, Diesmann, & Aertsen, 2002a, 2002b). Thus, the local and global tests can be considered complementary. In this letter, only local tests are considered.

8. **Analysis of temporal structure of templates**: If templates share a common temporal structure, this may have biological significance. Such structure can be easy to find in cases where neuronal activity is reliable and highly phasic (Dave & Margoliash, 2000), but it may be difficult to identify in cases where neuronal activity is variable and tonic. To deal with the variability in tonic spiking activity, an alignment procedure is developed from an information point of view. The idea is to minimize the uncertainty in predicting spiking events based on the empirical distributions induced by the temporal registries of the templates, hence forcing their common structure to be aligned. This approach is substantially different from the metric or cross-correlation-based methods, and it is based on considering the templates as a whole rather than pairwise comparisons (Victor & Purpura, 1996; Yu & Margoliash, 1996). In addition, the common temporal structure can provide useful cues for selection of filters.

2 Pattern Filtering

In this section, we develop pattern filtering for spiking activity. The same methodology can be established for general point processes.

2.1 Conditional Poisson Models for Spiking Activity. Since pattern filtering is based on a conditional Poisson assumption as described earlier, we need to specify the distributions of background sequences and target sequences.

For the background, the distribution is simply a Poisson process with a constant rate. To specify the distribution of target sequences, think of them as being generated by a template. Around each template spike, spikes are generated with random jitter. We assume that for all the template spikes, the spike-generating mechanism is the same, and therefore the distribution of the jitter is the same. In addition, "noise" spikes are generated outside the neighborhood of the template spikes. All these spikes considered together then consist of a spike train generated by the template.

More specifically, denote by $\mathcal{P}(f, A)$ a Poisson process on A with a rate function f. Let $S = \{s_1, \ldots, s_p\} \subset [0, \sigma]$ be a template and $I = [x, x + \sigma]$ be a time interval. A background sequence in I is simply a sample from $\mathcal{P}(q, I)$, with q a constant. On the other hand, to generate a sequence in I from S, first draw $Z_1, \ldots, Z_p$ independently and identically distributed (i.i.d.) $\sim \mathcal{P}(f_1, (-\epsilon, \epsilon))$ for a given rate function f_1 and $\epsilon > 0$. $Z_n + s_n$ is a sequence in $J_n := (s_n - \epsilon, s_n + \epsilon)$. We take it as the set of jittered spikes generated by s_n. The union of all J_n, $J = \bigcup_{n=1}^{p} J_n$, is the neighborhood of the entire template. Draw $X \sim \mathcal{P}(f_0, [0, \sigma]\backslash J)$, and take it as the set of noise spikes. Finally, "insert" $T_0 = X \cup \bigcup_{n=1}^{p}(Z_n + s_n)$ into the interval I. The resulting sequence $T = T_0 + x$ is a spike train generated by S in I. We will refer to a collection of sets X, Z_n as a "configuration" for T. The likelihood of the configurations, rather than the resulting spike trains, will be used in the classification.

The model incorporates the temporal detail of spiking activity around a template spike by f_1 and $\epsilon > 0$, with ϵ being the maximum jitter of a random spike generated by a template spike. In most cases, $f_1(x)$ is nonconstant, and we always assume it is nonincreasing in $|x|$ so that the probability decreases as the jitter increases. The above setup can be generalized to nonstationary noise within target spike trains by allowing f_0 to be a nonconstant. In this article, however, we will use constant f_0.

2.2 Classification by Likelihood Ratio and Pattern Filtering. Based on the above model, we now consider how to detect targets using one filter. Later we will discuss how to combine different filters. Target detection is achieved by classifying segments in a data spike sequence as either targets or background sequences. For a segment T, let $l_o(T)$ be the likelihood of T given that it is a background sequence, and $l_a(T)$ the *maximum likelihood* (ML) among all configurations for T, given that it is a target. Then the following log-likelihood ratio is a natural criteria for the classification,

$$L(T) = \log \frac{l_a(T)}{l_o(T)}.$$

Note that $l_a(T)$ is not the likelihood of T and often is smaller. As a consequence, L is not the difference between two rate functions. The choice of $l_a(T)$ reduces the chance of detecting targets that are not similar to the template. Also, it can be shown that $l_a(T)$ has a simpler form than the likelihood of T and hence is easier to compute.

To explicitly derive L and see how the detection across a long data sequence T can be accomplished by filtering, for each interval $[x, x + \sigma]$, let $T_x = T \cap [x, x + \sigma]$, where σ is the duration of a template $S = \{s_1, \ldots, s_p\}$. It is well known for Poisson processes that $l_o(T_x) = q^n \exp\{-q\sigma\}$. On the other hand, by the independence assumption in the model, $l_a(T_x) = Q_0 \times Q$, where Q_0 is the likelihood of the noise spikes in T_x and Q the ML of the jittered spikes generated by the template spikes. With $J = \bigcup_k (s_k - \epsilon, s_k + \epsilon)$, let the background function B and the time window function K be

$$B(x) = \begin{cases} \log(f_0(x)/q) & \text{if } x \notin J \\ 0 & \text{otherwise,} \end{cases}$$

$$K(x) = \begin{cases} \log(f_1(x)/q) & \text{if } x \in (-\epsilon, \epsilon) \\ 0 & \text{otherwise,} \end{cases} \tag{2.1}$$

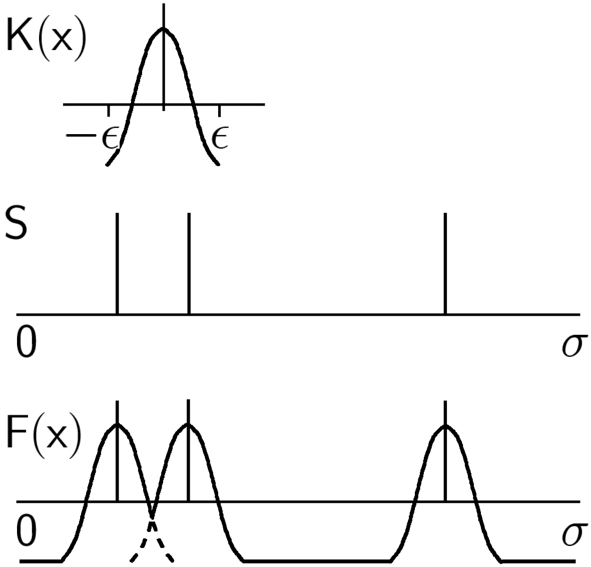

Figure 1: Construction of a filter. $K(x)$ is a function defined on $(-\epsilon, \epsilon)$. S is an exemplar spike sequence. In this case, it happens that $B(x) \equiv K(\epsilon) = K(-\epsilon)$. The filter F is plotted as the dark curve.

respectively, and $F(x) = F(x; S)$ as a function on $[0, \sigma]$, such that (see Figure 1),

$$F(x) = \begin{cases} \max_{s \in S} K(x - s) & \text{if } x \in J, \\ B(x) & \text{otherwise.} \end{cases} \tag{2.2}$$

Regard $T = \{t_n\}$ as a series of δ functions, that is, $T(t) = \sum \delta(t - t_n)$ (Rieke, Warland, de Ruyter van Steveninck, & Bialek, 1997). Then

$$L(T_x) = R(x) + C, \qquad \text{for any } x, \tag{2.3}$$

where C is a constant independent of S and x, and

$$R(x) = R(x; S, T) = \int_0^\sigma F(\tau)T(x + \tau)\,\mathrm{d}\tau.$$

Therefore, as x runs across T, up to a constant, $L(T_x)$ consists of a linear convolution of T. Thus, targets can be efficiently detected using filtering instead of classfication segment by segment. The proof of equation 2.3 is given in the appendix.

In practice, we select K and B by hand. We use local peaks in the filter response $R(x)$ to detect targets. Given prespecified $\theta > 0$ and $r > 0$, find all temporal points $x_1, \ldots, x_L$ satisfying

$$\text{(Local maximum)}\quad R(x_j) \geq R(x), \quad x \in (x_j - r, x_j + r) \tag{2.4}$$

$$\text{(Above threshold)}\quad R(x_j) \geq \theta. \tag{2.5}$$

The spike sequences in $[x_j, x_j + \sigma]$ are considered potential targets.

The filtering also applies to multiple point sequences, such as simultaneous activity of multiple neurons (Chi, 2003). Suppose $\boldsymbol{S} = (S_1, \ldots, S_p)$ is a set of spike sequences simultaneously recorded from p neurons in the time interval $[0, \sigma]$. Let $F_k(x) = F(x; S_k)$. Given a simultaneous recording $\boldsymbol{T} = (T_1(x), \ldots, T_p(x))$ from the same set of neurons, segments in $\boldsymbol{T}$ with spatiotemporal pattern similar to $\boldsymbol{S}$ are located by equations 2.4 and 2.5, with

$$R(x) = \sum\nolimits_{k=1}^{p} \int_0^{\sigma} F_k(x) T_k(x + \tau)\, \mathrm{d}\tau.$$

It can be shown that both the sliding sweeps algorithm and cross-correlation are special cases of linear convolution. One can choose $K(x) = 1$ and $B(x) = -M$ for large $M > 0$ to achieve the same effect as the sliding sweeps algorithm. Cross-correlation is essentially convolution of the rectified spike counts of the data sequence with that of the template.

3 Multiple Pattern Filtering Procedures

As we argue in section 1, to improve the detection of spiking activity, multiple pattern filters, templates, and temporal scales can be incorporated. Figure 2 illustrates how the detection process proceeds. For each template S_n, every segment of the data sequence T is matched with it using several types of filters, one after another. If the response of the segment to one of the filters is subthreshold, it is classified as a background sequence and will not be treated by subsequent filters. To account for time scaling in the activity, the segment is matched with the template at different time scales. The outputs associated with the template, therefore, are segments that match the template at some time scale. Among the segments, only those with small p-values under a test P at the end of the channel are output as targets. Finally, because a segment may be detected multiple times, by filters associated with different templates, the target outputs from the channels are combined, so that only one is chosen from overlapping targets.

Throughout the analysis, detection, training, and significance testing follow the same protocols. For all the templates, the associated filters are constructed with the same time windows and background functions and are scaled by the same set of factors. Likewise, training for different templates results in different parameter values. The variability of the templates thus is incorporated throughout the detection process. This makes the process more robust to variability, a desirable property for pattern detection.

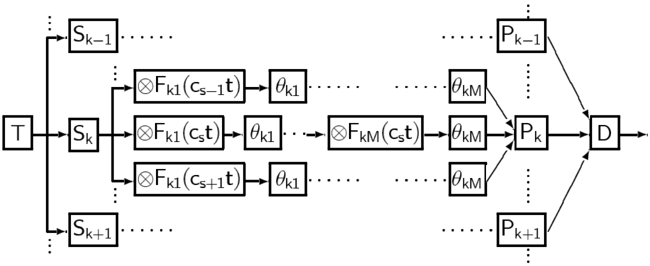

Figure 2: Detection by multiple templates, filters, at multiple scales. T is a data spike sequence. S_k are templates, for all of which the same set of window and background functions are used to contruct pattern filters. F_{km} are filters constructed from S_k, using the mth window function and background function. θ_{km} are associated thresholds. Only segments in T with above-threshold response to F_{km} are considered potential targets. c_s are scaling factors applied to all S_k. P_k are significance tests on targets detected by filters for S_k. Only targets with high significance are kept. To avoid overcounting, D is a procedure to choose only one target from each set of overlapping targets detected across the channels.

3.1 Detection at Multiple Levels

3.1.1 Multiple Filter Types. A pattern filter enforces some constraints on the statistical assumptions regarding spiking activity. As long as the filter is constructed based on empirical evidence, the constraints are biologically plausible. Multiple filters induce more constraints, which may lead to a more realistic probability model of spiking activity and, hence, better detection. The resulting model is only implicitly specified; however, it allows for efficient computation for detection.

Given a template, using different time window and background functions, multiple filters can be constructed from the template. The square window function,

$$K(x) = \mathbf{1}_{(-\epsilon,\epsilon)}(x) = \begin{cases} 1 & \text{if } x \in (-\epsilon, \epsilon) \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

and $B(x) \equiv 0$ are among the simplest choice. One may also apply other time window and background functions for a filter, such as

$$K(x) = \begin{cases} \frac{1}{2}(1-\beta) + \frac{1}{2}(1+\beta)\cos\frac{\pi x}{\epsilon}, & \text{if } x \in (-\epsilon, \epsilon) \\ -\beta & \text{otherwise} \end{cases},$$

$$B(x) \equiv -\beta, \quad \beta \geq 0. \tag{3.2}$$

K essentially is a Hamming window. Filters made from equation 3.2 are continuous and nonconstant around template spikes. Another reasonable choice includes "cropped" versions of K,

$$\bar{K}(x) = \min(\alpha, K(x)), \quad \alpha \in (0, 1), \tag{3.3}$$

which can be thought of as mixtures of square windows and Hamming windows.

Computational efficiency is maintained if the filters are applied in a coarse-to-fine manner (see Figure 2). To detect targets, the first filter is applied to the entire data spike sequence to detect segments with bare similarity to the template. In actual computation, this step is fast, requiring convolving the filter and the data, both discretized only at a coarse resolution. Subsequent filters are used only for the detected segments. In these steps, the filter and the segments are discretized at finer resolution. Any segment that yields a below-threshold response is classified as background and is excluded. Since the total duration of the segments is much shorter than the entire sequence, the subsequent steps of filtering are also fast.

3.1.2 Multiple Scales. Spiking activity often exhibits across-trial nonstationarity, which is different from across-time nonstationarity (Chi & Margoliash, 2001; Grün et al., 2002a). A remarkable type of across-trial nonstationarity is temporal scaling, which is observed across different states of the brain (Nádasdy et al., 1999; Dave & Margoliash, 2000; Louie & Wilson, 2001). To take this into account in the detection process, we use filters scaled in time at several levels.

To detect targets in T at a time scale c, for each filter F, use $R_c(x) = \int_0^\sigma F(c\tau; S)T(x+\tau)\,d\tau$ in place of equation 2.3. To account for the scaling, the radius r in equation 2.4 is changed to cr, while the threshold θ in equation 2.5 remains the same.

The coarse-to-fine multiscale procedure is summarized as follows. Given a template of duration σ, suppose M filters are applied, with corresponding thresholds $\theta_1, \ldots, \theta_M$, and the detection is conducted at D time scales. Then, for each scale c_n,

```
Filter $T$ by $F_1(c_n t)$ to get $R_1$
$L_n \leftarrow \{x : R_1(x) \geq \theta_1,$ and $R_1(x) \geq R_1(t),$
          for any $t \in (x - c_n r, x + c_n r)\}$
for $m = 2, \ldots, M$                                              (3.4)
  for $x \in L_n$
    Filter $T \cap [x - c_n r, x + c_n \sigma + c_n r]$ by $F_m(c_n t)$ to get $R_m$
    Remove $x$ from $L_n$ if $R_m(t) < \theta_m$ for all $t$
Output $L_n$
```

3.1.3 Multiple Templates. The conditional Poisson model describes the variability of spike sequences resulting from small random perturbations of a template. On a more global scale, there can be variability across the templates that is not accounted for by the Poisson model. When this happens, multiple templates should be incorporated.

Suppose M pairs of time window and background functions (K_m, B_m) are given. Then for template S_n, $n = 1, \ldots, N$, we repeat the detection, equation 3.4, using filters $F_m^{(n)}$ constructed from S_n and (K_m, B_m). The target sets associated with different templates are combined (see section 3.4). For each template, the thresholds are set relatively high, so only a small number of segments can be detected. Thus, the detection for each template is quite specific. However, with many templates being incorporated, the overall detection covers a reasonably wide range of target activity.

Detection is faster using mean filters $\tilde{F}_m = \frac{1}{N}\sum_{n=1}^{N} F_m^{(n)}$ for each pair (K_m, B_m). However, since the average filters do not represent the variability of the templates, this approach has the same drawback as the detection using only a single template. Thus, we do not advocate using mean filters.

3.2 Training. The goal of training is to select thresholds for filter responses in order to calibrate the detection. For each filter, the threshold depends on the responses of two training sets: one consisting of background activity and the other target activity. In principle, both should be sampled under similar conditions (i.e., behavioral states); for example, if the filter is meant to detect replayed patterns in spontaneous activity during sleep, the training data should consist of such activity. The problem is that unlike in supervised learning, often there is no reliable way to tag the sample activity as background or targets. As an alternative, simulated sequences are substituted as training data.

To simulate the responses of background activity to a filter F, sequences sampled from a homogeneous Poisson process of the same duration as the corresponding template are convolved with F. The rate of the process is the average firing rate of a long neuronal trace recorded from the same state as the data spike sequence, which is reasonable when targets are rare events. For each sample, register the peak value of its convolution with F. Then let $\alpha(F) =$ the 99th percentile of the peak values.

To simulate the responses of target activity to F, randomly modified templates are used as simulated targets. Given a template S, a simulated target is generated in four steps:

1. Random deletion of each spike in S with a small probability.
2. Random shift of each remaining spike in S by distance $x \sim N(0, \delta)$.
3. Add a sample from a Poisson process with low density to S as noise spikes.
4. Random scaling of S by $e^{-\xi}$, with $\xi \sim \text{Uniform}(-\epsilon, \epsilon)$.

Given a percentile $\pi \geq 50$ only depending on the time window and background functions used by F, let $\mathbf{S_F}$ be the set of the peak responses to

F of the simulated targets. Let $\beta(F)$ = the πth percentile of $\mathbf{S_F}$. Then the threshold for F is chosen to be

$$\theta = \theta(F) = \max\{\alpha(F), \beta(F)\}. \tag{3.5}$$

This threshold is high enough to exclude most of the background activity and to ensure that the detected targets are similar to the template.

3.3 Significance Test. It is often the case that the average firing rate around a target is significantly different from the average firing rate across the data sequence. This raises the possibility that some targets might be artifacts generated by short episodes, which have constant but significantly different firing rates from their surroundings. This possibility cannot be ruled out by pattern filtering. To control for it, some statistical test of the significance of the targets is needed.

Suppose the sequence in $[t_0, t_1]$ is detected. To test whether the target is just a segment of constant rate activity, the following procedure is conducted,

1. Calculate the average firing rate r within an interval J containing $[t_0, t_1]$.
2. Sample N sequences from a Poisson process with density r on J, with $N \gg 1$.
3. For each sampled sequence, repeat the same detection procedure across all timescales.
4. Count the number n of sequences in which targets are detected at any of the scales.
5. Output $p = n/N$ as the p-value under the null hypothesis.

To account for the physiological limit of a neuron, in actual implementation, before step 3, spikes with temporal distance from preceding spikes less than a certain value are deleted from the sampled sequences. The modified sequences are not random samples from a Poisson process anymore, but still are derived from a process with a constant firing rate.

3.4 Ambiguity Solving. If the templates are exactly the same, then at a given temporal scale, the detected targets should be the same across the templates. Due to the variability of the templates, targets detected with different templates often overlap in time but are not exactly the same. Thus, a short time interval can be associated with multiple targets, causing ambiguity of detection. Detection at similar temporal scales may also lead to ambiguity.

To avoid overcounting, it is sensible to choose only one from overlapping targets. A reasonable criterion is the p-value of a target obtained from the test in section 3.3, so that among overlapping targets, only those with the smallest p-value are kept. When two or more targets remain, one can adopt some ad hoc criteria to choose between them.

4 Information-Based Alignment of Samples

Given a sample of spike sequences, usually associated with the same stimulus or behavior, temporal alignment refers to adjustment of the temporal registry of spikes under certain constraints, so that a characteristic common temporal pattern of the spike sequences can be revealed, which otherwise is difficult to observe (see Figure 6).

There are several alignment methods. One is based on dynamical programming, which combines aspects of rate coding and temporal coding (Victor & Purpura, 1996). In this method, there are no "hard" constraints on the alignment. Instead, any adjustment of the temporal registry is allowed but with a certain cost. The optimal adjustment is the one that minimizes the total cost. Another method cross-correlates the spike sequences or the associated behaviors (such as spectrograms of vocalizations) and registers the spikes when the cross-correlation reaches maximum (Yu & Margoliash, 1996). In this method, the ISIs of the spike sequences are kept constant, and the only way to adjust the temporal registry of the spikes is rigid shifting of the entire sequences. Furthermore, both methods make pair-wise comparisons of templates. This leads to a series of local optimum, which does not guarantee achieving a global optimum. A third method aligns templates by minimizing a global measure of distance. The method works well when the spiking activity is highly phasic (Chi & Margoliash, 2001). However, it is not as well suited for tonic activity.

The above alignment methods are based on the idea of reducing certain measures of distance among spike sequences. The alignment introduced here is from a different perspective. To illustrate this, we consider the following problem. Suppose we have the temporal registries of spike sequences $S_1, \ldots, S_N$, and we find that at each time t, a certain event is observed in some of the spike sequences but not in the others. Assume these spike sequences are collected under the same conditions. One may ask, If we are to collect another spike sequence S under the same conditions, then, based on $S_1, \ldots, S_N$, how well can we predict that the same event will occur in S at t?

Intuitively, at any time where common temporal structure among S_n occurs, which can be excitation, inhibition, or something else, an event will have a distribution concentrated around 1 or 0. Based on the distribution, the uncertainty in predicting the event at that time is low. Consequently, the overall uncertainty in predicting events across time is good indication of how well the common temporal structure of S_n is expressed across time, which clearly depends on the temporal registries of S_n. The lower the uncertainty is, the more explicitly the common structure is expressed. In information theory, uncertainty is measured by entropy. The alignment can thus be achieved by reducing the total entropy of the events across time.

In actual computation, to deal with the problem of sparse data, it is important to choose events appropriately. For each time t, we define

$$A_t = \{\text{At time } t\text{, spiking occurs in } [t - \tfrac{\Delta}{2}, t + \tfrac{\Delta}{2})\},$$
$$X_t = X_t(S) = \begin{cases} 1 & \text{if } A_t \text{ happens to } S \\ 0 & \text{otherwise} \end{cases}. \tag{4.1}$$

Note that A_t does not involve the exact number of spikes. The advantage of using equation 4.1 is that X_t is a Bernoulli random variable, so the estimation of $P(X_t = 1)$ does not require a large number of templates. The estimate of $P(X_t = 1)$ is

$$\hat{P}_t = \hat{P}_t(S_1, \ldots, S_N) = \frac{\text{Number of } S_i \text{ having spikes in } [t - \frac{\Delta}{2}, t + \frac{\Delta}{2}]}{N}$$
$$= \frac{1}{N}\sum_{n=1}^{N} X_t(S_n), \tag{4.2}$$

with estimated std $\sqrt{\hat{P}_t(1 - \hat{P}_t)/N - 1}$. The entropy of the Bernoulli random variable is equal to

$$\hat{h}_t = -\hat{P}_t \ln \hat{P}_t - (1 - \hat{P}_t)\ln(1 - \hat{P}_t), \tag{4.3}$$

with $0 \ln 0$ defined to be 0. Define

$$H(S_1, \ldots, S_N) = \int \hat{h}_t \, \mathrm{d}t, \tag{4.4}$$

which is an upper bound of the joint entropy of the process X_t, $t \in \mathbf{R}$ (theorem 2.6.6, Cover & Thomas, 1991). The information-based alignment of $S_1, \ldots, S_N$ is to find $t_1, \ldots, t_N$, such that

$$H(S_1 + t_1, \ldots, S_N + t_N)$$
$$= \min\{H(S_1 + \tau_1, \ldots, S_N + \tau_N) : \tau_1, \ldots, \tau_N \in (-\infty, \infty)\}. \tag{4.5}$$

For actual data, $H(S_1 + \tau_1, \ldots, S_N + \tau_N)$ is a complex function in $\{\tau_n\}$. It can be minimized by stochastic annealing (Geman & Geman, 1984) rather slowly. Alternatively, we approximate its minimum by a randomized greedy procedure. For each cycle, in a random order, we shift S_n, one at a time, to minimize $g_n(t) = H(S_1, \ldots, S_n + t, \ldots, S_N)$ and update S_n to $S_n + t$ if t minimizes g_n. We then carry out multiple such cycles of minimization, choosing a different random order at each cycle, until $H(S_1, \ldots, S_N)$ cannot be reduced by shifting individual S_n.

Updating S_n in a random order reduces the possibility of settling into a local minimum. Experience indicates that many cycles are not required. Usually 10 to 15 cycles suffice. Although it is possible to reduce the entropy further by annealing, as a practical matter for the data we analyzed, this did not seem to appreciably improve the alignment and was not implemented.

We now remark on some features of the information-based alignment. The most significant difference between the information-based alignment as compared to the previous pairwise alignment procedure is that information-based alignment is based on the templates as a whole, which naturally allows an information-theoretic interpretation. In contrast, pairwise alignment puts emphasis on the common structure between pairs of sequences, which may not be the common structure of all the sequences. This leads to a series of local minima. One can imagine that the firing-rate function induced from aligned templates should be "clumpy" instead of flat. Entropy is an objective measure of the clumpiness.

Second, the definition of the event A_t involves a parameter Δ. If Δ is too large, then by equation 4.2, $\hat{P}_t \approx 1$ for most t over the time inteval of $S_1, \ldots, S_N$, and hence $H(S_1, \ldots, S_N) \approx 0$. On the other hand, if Δ is too small, because of the variabily in spike timing, no matter how the temporal registries of $S_1, \ldots, S_N$ are adjusted, for most t, the number of S_i with spikes in $[t - \frac{\Delta}{2}, t + \frac{\Delta}{2}]$ is at most 1. It is then not hard to see that $H(S_1, \ldots, S_N) \approx CN \log N$, where C is the total number of spikes in $S_1, \ldots, S_N$. In either case, the associated entropy is difficult to reduce, and hence $\{S_n\}$ will not be aligned. For HVc activity, we got satisfactory result when Δ^{-1} was at the same order of the maximum firing rate. Presumably, optimal Δ depends on the intrinsic temporal precision of spiking activity (Victor & Purpura, 1996). This is subject to further study.

Third, in the optimization, equation 4.5, only rigid shifts are allowed for S_n. The reason is twofold. Since the internal structure of S_n is maintained, the alignment can reveal systematic change in spiking activity that otherwise would not be revealed by shifting individual parts (Chi & Margoliash, 2001). Presumably, by taking into account time scaling, alignment of temporally morphed templates may yield better results. However, if all the templates are collected under the same conditions, there is no compelling argument why morphing of the templates is more reasonable than not doing so. Nevertheless, depending on problem at hand, even with rigid shift, individual parts of S_n can be shifted relative to the others (Yu & Margoliash, 1996).

5 Experiments

We tested pattern filtering on neuronal data collected from the HVc of the zebra finch (*Taeniopygia guttata*). The song system is a standard object in the study of neural mechanisms of vocal learning, production, and maintenance (Brenowitz, Margoliash, & Nordeen, 1997). HVc is a nucleus in the forebrain of the birdsong system. It plays an important role in vocal learning, auditory input integration, and higher-level motor commands for vocalization. HVc directly projects to another forebrain nucleus, robustus archistriatalis (RA). Recently, it was found that in sleeping zebra finches, the spontaneous spiking activity of RA neurons occasionally exhibited spike bursting patterns similar to the premotor activity patterns that the same neurons exhibit during singing and in response to auditory playback of the bird's own song (BOS) during sleep (Dave, Yu, & Margoliash, 1998; Dave & Margoliash, 2000). This "replay phenomenon" during sleep is hypothesized to play an important role in learning and memory consolidation of the birdsong system (Margoliash, 2001). Since RA receives inputs from HVc, it is natural to consider whether the replay phenomenon also occurs in the spontaneous activity of HVc and is potentially the source of the replay in the spontaneous activity of RA. Recent recordings in sleep-induced birds from RA-projecting HVc neurons support this interpretation (Hahnloser, Kozhevnikov, & Fee, 2002). Furthermore, the activity in HVc interneurons, which are the presumed class of neurons analyzed in this article, has far more variability than the activity in RA (Mooney, 2000; Shea, Rauske, & Margoliash, 2001), and therefore the application of pattern filtering to HVc data is both biologically interesting and a useful test of the effectiveness of the techniques developed in previous sections.

The steps of the detection are (1) template selection, (2) training of the detector as described in section 3.2, and (3) the multiple detection procedures in section 3.1.

During sleep, HVc auditory responses to BOS often exhibit considerable cross-trial variability. Some unobserved variable, such as phase of intrinsic bursting within the song system or different stages of sleep, presumably is related to the variability of auditory responses. From the raster plot in Figure 3, it can be seen that in many trials, the auditory responses were quite weak and less structured. To achieve effective detection, sample sequences associated with robust responses and clearer structure were chosen. First, sequences without enough spikes (< 20) were removed from the sample. Templates were then chosen from the remaining sequence. In general, if a sequence did not match well with the other sequences—for example, on average there were too many nonmatching spikes—then it was not chosen as a template.

5.1 Results. We analyzed the sleep activity of seven single units in HVc of four zebra finches (two units per bird for three birds and one unit for one bird; see Table 1). For each unit, the data consisted of activity that was collected in a large number of trials while the animal slept. At the beginning of each trial, a recording of BOS was broadcast to the animal, and neuronal responses to the stimulus were collected. The spontaneous activ-

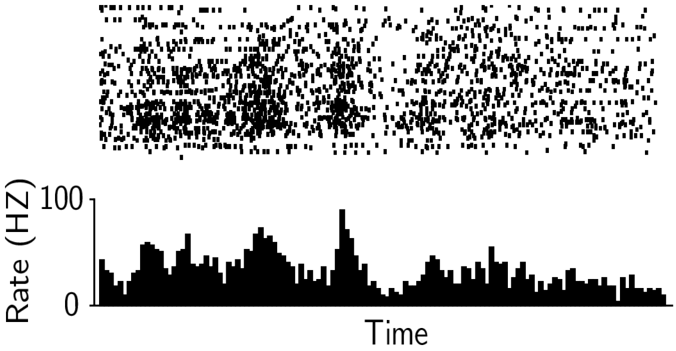

Figure 3: Auditory responses of HVc neuron unit 1 to multiple repetitions of the same motif in BOS. (Top) Raster plot of the spike sequences of auditory responses, starting at the onset of the motif. The spike sequences are aligned at the onset. The duration of each spike sequence is about 500 ms. (Bottom) The firing-rate function of the spikes sequences computed using a disjoint time window of size 5 ms (Dayan & Abbott, 2001).

Table 1: Results of Detection.

Bird	Unit	*T* (ms)	*N*	*n*
hv31_1005	1	504 ± 5	5	0.68
	2	504 ± 5	30	4.08
hv33_1106	3	620 ± 0	31	1.27
	4	570 ± 0	21	0.86
hv39_1014	5	659 ± 4	21	0.46
	6	661 ± 4	50	3
hv40_0626	7	665 ± 3	3	0.82

Notes: *T*: Duration of template. *N*: Number of detected spike sequences. *n*: Average number of detected sequences per minute.

ity of the neuron was collected in the intervals between presentation of BOS (approximately once every 20 ± 6.5 s). The purpose of the experiments was to investigate if auditory responses were replayed during spontaneous activity of sleep in HVc.

For each bird, the recording of BOS had one or two renditions of a motif (sequence of repeated syllables). Templates were selected from the auditory responses of the HVc units to the renditions. Template duration *T* is defined as the duration of the corresponding stimulus. For each unit, we chose one time interval during which its responses stayed relatively strong across the trials and selected templates from those responses. Thus, mean(*T*) can differ for units of the same bird.

Table 1 lists the total numbers and temporal densities of detected spike sequences from spontaneous activity that exhibited similar temporal patterns as the templates. Across the units, filters constructed with the following three types of window functions and background functions were applied sequentially: square window (see equation 3.1), Hamming-type window (see equation 3.2), and cropped versions of Hamming windows (see equation 3.3). The same set of parameter values for pattern filters was used for the results reported in Table 1. It is noteworthy that for each of the units analyzed some spontaneous patterns were found to match templates drawn from auditory responses to BOS.

Next, we show the results for two units in more detail, each involving a single HVc neuron of a different zebra finch.

5.1.1 Unit 1. The BOS consisted of two renditions of a motif and was broadcast in 63 trials, leading to 126 spike sequences of auditory responses to the playbacks of the motif. Figure 3 displays all the spike sequences, which are aligned by the onset of the motif.

Of the 126 spike sequences, 48 were chosen as templates. Figure 4A shows the raster plot and estimated firing-rate functions for these trials. The templates were then aligned by minimizing entropy. The raster plot and the firing-rate function of the aligned templates are shown in Figure 4B. The raster plot of the aligned templates clearly reveals more temporal structure than for the onset-registered templates. The standard error of the shifts of the aligned templates was about 5 ms, with the difference between the maximum shift and the minimum shift equal to 30 ms. The average shift is not important, because only the relative shifts of the templates matter for the alignment.

To reduce the amount of computation, 30 of the 48 spike sequences are further selected. The constructed pattern filters were scaled at five discrete levels, $c = 1, 0.8, 0.9, 1.1$, and 1.25, and convolved with the data spike sequences following, as described in section 3.1.

For each aligned template, detection was conducted for the recorded spontaneous activity across all 63 trials, each lasting 7 seconds. A total of 22 spike sequences, which yielded above-threshold responses to all the filters, were found. Among them, 8 were highly significant ($p \leq 0.01$) by the test described in section 3.3. After disambiguation, only 5 remained and were output as targets. In Figure 5, all of the targets are displayed along with the templates used to detect them. In order to be compared with the templates, the targets were scaled in time at the levels at which they were detected. The variability among the templates is obvious, which argues for detection based on individual templates instead of some form of average of the templates.

5.1.2 Unit 3. Unlike the previous case, the BOS for this bird contained only one rendition of a motif lasting about 620 ms and was broadcast to

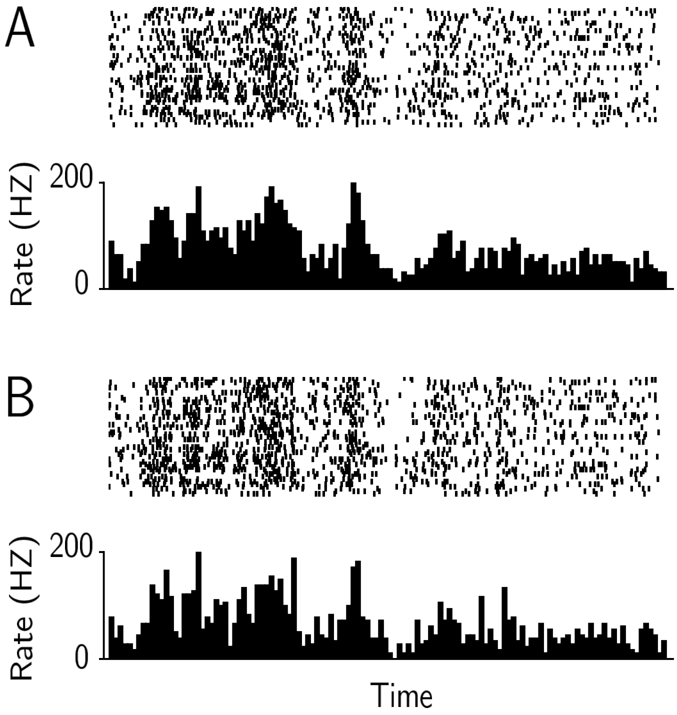

Figure 4: Raster plot and empirical firing rate function of selected sample sequences from unit 1, before alignment (A) and after alignment (B). The firing rates were estimated by the same method as in Figure 3. Note the different scales of the firing-rate functions in *A*, *B*, and Figure 3.

the sleeping bird in 126 trials. Figure 6A is a raster plot of the 126 spike sequences of auditory responses to the motif, aligned by the onset of the playback of the motif. There is an obvious change in the latency of the responses. Of the 126 spike sequences, 76 were selected as templates. On average, there were 29.4 spikes in each template. Figures 6B and 6C are the raster plots of the selected sequences before and after they were aligned by minimizing entropy. As for the previous neuron, the information-based alignment reveals some noticeable common patterns of the responses.

The data of spontaneous activity during sleep were about 25 minutes in total duration. Using 30 of the 76 selected sequences as templates, 31 nonoverlapping targets with p-value ≤ 0.01 were detected. The relatively large number of detected targets allowed us to have a more global view of the temporal pattern of the spontaneous activity, as compared to the auditory responses. First, for each target detected at scale c, its scaled version by factor $1/c$ was generated. Then the scaled targets were combined into a raster plot,

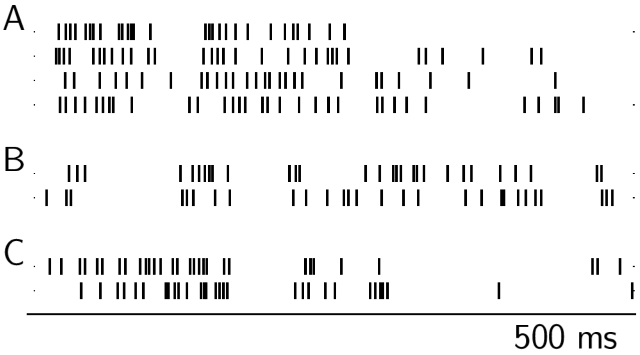

Figure 5: Templates and targets (unit 1). In each group, the first trace is a template; the others are targets detected with filters built from the template. The targets are scaled in time in order to be compared with the templates (in *A*, by $\frac{1}{0.9}$, $\frac{1}{1.25}$, $\frac{1}{1.25}$, respectively; in *B*, by $\frac{1}{1.1}$; in *C*, by $\frac{1}{1.25}$). The line at the bottom represents an interval of 500 ms. No spikes occurred in the blank regions in the traces.

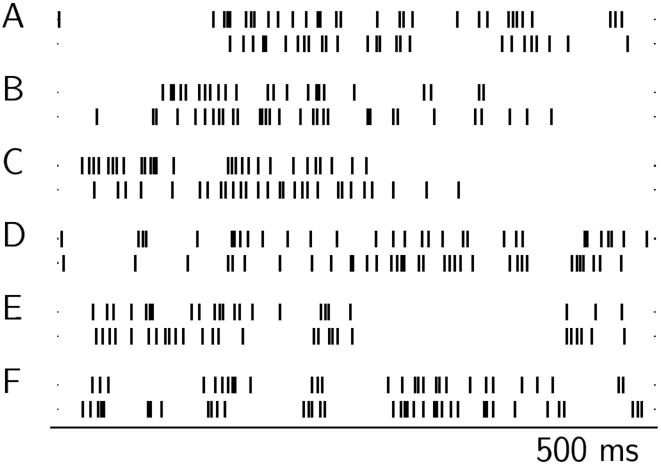

Figure 6: (A) Raster plot of 126 sample spike sequences of HVc neuron unit 3 in response to playbacks of the BOS. All the sequences are about 620 ms long and aligned at the onset of the BOS. (B) Raster plot of 76 selected sequences used as templates. The order of the sequences is the same as in the original sample. (C) Raster plot of the same 76 spike sequences, after they are aligned by entropy minimization. (D) Raster plot of 31 spike sequences in spontaneous activity of the same neuron during sleep. Among them, 1 is scaled by $\frac{1}{0.9}$, 7 by $\frac{1}{1.1}$, and 13 by $\frac{1}{1.25}$. The others are unscaled.

with those detected in the earlier part of the spontaneous activity plotted on top of those detected in the later part. The raster plot is shown in Figure 6D, which clearly demonstrates the similarity between the responses to BOS and the spontaneous spike sequences.

5.2 Sensitivity to Parameters. To test how different values of the parameters affect the detection, we compared the results of the detection by changing the value of one of the parameters while fixing the others. This required the entire detection procedure, including training and the significance test, to be rerun.

The main parameters are those for the time window and background functions, specifically,

- The window sizes ϵ_1 for the square window (see equation 3.1), ϵ_2 for the Hamming window (see equation 3.2), and ϵ_3 for the cropped Hamming window (see equation 3.3)
- The background values β_1, β_2, β_3, respectively, associated with the above window functions to construct the filters.

Because of the similarity between the Hamming window function and its cropped versions, we always set $\epsilon_2 = \epsilon_3$ and $\beta_2 = \beta_3$. In both cases reported in the previous two sections, the values of the parameters were $\epsilon_1 = 4$ ms, $\epsilon_2 = \epsilon_3 = 5$ ms, $\beta_1 = -0.3$, and $\beta_2 = \beta_3 = -0.4$.

Suppose targets in time intervals J_k, $k = 1, \ldots, K$ were detected with the above parameter values. To analyze the targets detected when one of the parameter values was changed, they were classified into three categories. A target detected in interval I is classified as new if $I \cap J_k = \emptyset$ for all k. It is classified as already detected if it has a significant large overlap with some J_k. Specifically, the overlap between I and J_k covers more than $c = 4/5$ of J_k. Otherwise, it is classified as overlapping. In general, if the number of already detected targets is large, then it indicates that different parameters lead to consistent detection. The larger c is, the harder it is to classify a target as already detected. We also tested $c = 9/10$, and the results were very similar.

For the first HVc neuron analyzed here (unit 1), we first doubled ϵ_1 to 8 ms while keeping the other parameters unchanged. Following the same procedures for training, detection, and significance test, seven targets were detected. Among them, two had been previously detected, while the other five were new, shown in Figures 7A through 7E. When ϵ_1 was reset to 4 ms, while ϵ_2, β_1, or β_2 was doubled, all but one target were already detected either in Figures 5 or 7A through 7E. The new target is displayed in Figure 7F.

We applied the same procedure to data for unit 3. Table 2 collects the results. Each column summarizes to the detection with the value of the corresponding parameter being doubled, while keeping the others fixed.

The results show that different parameter values have some, but not a large, impact on detection. On the one hand, new targets may be detected.

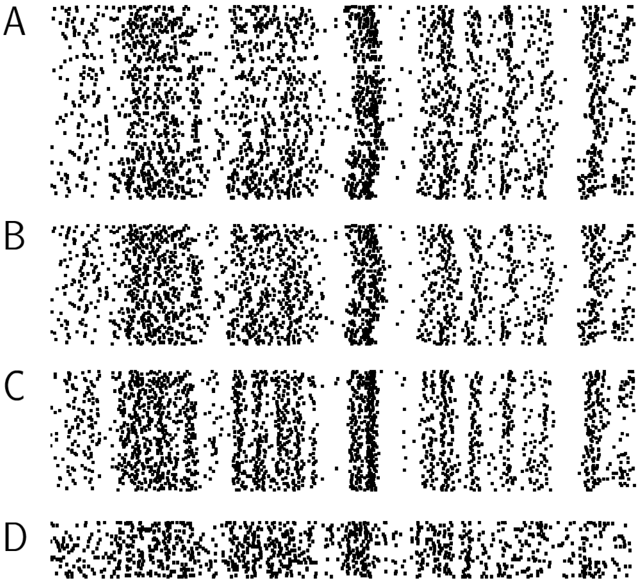

Figure 7: Templates and corresponding targets detected from the same data as Figure 5, but with the window size of one of the time window functions doubled (see section 5.2). The targets are scaled by $\frac{1}{1.1}$, $\frac{1}{1.25}$, $\frac{1}{1.1}$, $\frac{1}{1.25}$, 1, and $\frac{1}{1.31}$, respectively. As a window size increases, more jitter is allowed in detection, leading to less close matching of some of the detected targets and the corresponding templates.

Table 2: Breakdown of Targets Detected for the Second HVc Neuron by Doubling the Value of One Parameter, as Compared with Targets Detected When $\epsilon_1 = 4$ ms, $\epsilon_2 = \epsilon_3 = 5$ ms, $\beta_1 = -0.3$, and $\beta_2 = \beta_3 = -0.4$.

	ϵ_1	ϵ_2	β_1	β_2
Already detected	18	25	16	14
Overlapping	5	4	4	1
New	5	7	5	5
Total	28	36	25	20

On the other, when the number of targets is relatively large, most of the targets detected using one set of parameters can also be detected using different sets of parameters, up to a relatively small difference in time. This is not unexpected, because different window sizes correspond to different statistical models on spiking activity. Furthermore, window sizes also affect the training and significance test on the targets. These changes, when combined, lead to different detection results. It is thus a good strategy to try different parameter values that are biologically plausible and report targets that are significant under a specific statistical test.

6 Discussion

We have described a novel approach to spike sequence detection: pattern filtering. Under the framework of temporal point processes, the approach incorporates (1) conditional Poisson processes to describe spiking activity, (2) detection by classification based on likelihood ratio, (3) classification by filtering, (4) multiple levels of detection, and (5) information-based temporal alignment of spike trains. Pattern filtering has significant computational advantages. First, it effectively detects spike sequences with a specific temporal pattern. Second, it allows the training and significance testing of the detection to be done efficiently, which otherwise would normally involve substantial computation. To improve pattern filtering, there are several issues to be addressed.

6.1 Statistical Significance of Targets. When a target is detected, it is necessary to make sure that it is not likely to have resulted by chance from background activity. One thus needs to test the *p*-value of the target: the likelihood of the target if it was generated by the background. Unfortunately, the probability distribution of the background activity is often poorly understood, making theoretical evaluation of the *p*-value infeasible.

The approach we take to address the problem is to use local processes as surrogates for the background activity (see section 3.3). The processes are modified homogeneous Poisson with short ISIs being removed to account for the neuronal refractory period. It is well known that a homogeneous Poisson process in general is not a good model for spiking activity (Riehle et al., 1997). At least, it does not account for rate modulations and possible nonstationarity of the spiking activity. This is a reason for using local processes, which can approximate the probability distributions around targets better than a global, homogeneous Poisson model does. For HVc, this choice of surrogates seems to work reasonably well. For other systems, however, different surrogate processes may be more appropriate. For example, in cases where neurons exhibit very low-frequency discharge but occasionally burst strongly, a Poisson distribution will not fit the data well, and a different distribution should be considered. In any case, localized processes will not only be sufficient, but will be better than a global surrogate process for the *p*-value tests of individual targets.

The detection considered here does not address another aspect of the statistical significance of targets: their global properties, such as the total number of targets. This has been a challenge for quite a long time, because it requires modeling of the entire data set (Abeles & Gerstein, 1988; Abeles et al., 1993; Date et al., 1998; Baker & Lemon, 2000; Grün et al., 2002a). In the light of the above comments, one possible approach to the global statistical significance of targets is to use a hierarchy of processes. Intuitively, this means that for each data segment, which may have random onset and offset, we use a local process to model it, and the process at the top of the hierarchy specifies what type of process to use for the segment as well as when the segment starts and ends.

6.2 Multiple Filtering. The basic procedure in our multilevel detection is serial (see Figure 2), as the filters in each channel are arranged in the form $F_1 \to F_2 \to \cdots \to F_n$, with F_1 having a coarser temporal resolution (i.e., sampling rate) than the others. This feedforward arrangement of filters can be changed into a parallel one, whereby each F_k outputs to a common integrator G independent of the others. Since the coarser temporal resolution of F_1 makes it faster to compute its output, G will first check the output of F_1, and only when the output is above threshold will G check the outputs of the other F_k. If above-threshold outputs from $F_1, \ldots, F_n$ coincide at a location x in the data, then G outputs the data at x as a target. This parallel-feedforward mechanism is biologically plausible, because it keeps the sensors in a system operating on a continuous basis, without being interfered with by the others at the same level of hierarchy.

6.3 Estimation of Density Functions for Filters. In our study, the time window functions and the background functions were chosen by hand. It is possible to estimate both from a sample of spike sequences, which potentially may lead to a better probability model of the spiking activity.

First, from the discussion in section 2.2, both the time window and background functions of a filter depend on the average firing rate at background, which is easily estimated from data. From a theoretical point of view, the underlying assumption that the background spiking activity follows a homogeneous Poisson process over time is questionable. One possible modification is to divide the background activity into several categories and model each one with a Poisson process, which may be inhomogeneous. In the appendix, we suggest a non-Bayesian approach as well as a simple Bayesian approach to incorporate different categories of background. It is worth mentioning, however, that for the detection per se, the homogeneous Poisson model of background activity has worked well.

Second, the density functions of jittered spikes around template spikes and noise spikes within a target can also be estimated. To pursue this, temporal alignment will play an important role. The appendix suggests a fairly simple but crude estimation procedure for the density functions. However, the point is that the estimation of the density functions requires the distinction between jittered spikes and noise spikes, which is hard to made without appropriate alignment.

6.4 Alignment. Our study demonstrates that alignment of spike sequences is useful for understanding the temporal discharge patterns of higher-order neurons such as in HVc. The responses of such neurons may exhibit internal structure yet variability in phase in relation to the stimulus. This variability may manifest itself as a systematic change in response latency across consecutive trials, as exhibited in Figure 6. Besides systematic modulations, there can also be random or local perturbations of the system, resulting in magnified random changes in spike timing across trials.

Basically, the alignment aims to "undo" either the systematic modulations or the random changes in the spike sequences. The information-based alignment is meant for tonic neuronal activity, where the notion of distance is harder to define. The alignment takes into account only the entropy of a binary spiking event in each individual time bin and then sums up the entropy. Note that the binary spiking event is not directly related to the exact number of spikes. Indeed, the binary events can be considered quantization of the actual spiking events. Simple spiking events such as binary events can be used for alignment and require far fewer samples.

In the alignment considered here, any correlation among spikes in different time bins is ignored by this alignment. One can indeed use joint binary events. For example, with X_t defined as in equation 4.1, given time lags $\Delta_1, \ldots, \Delta_{n-1}$, let $Y_t = (X_t, X_{t+\Delta_1}, \ldots, X_{t+\Delta_{n-1}})$. Then, in equation 4.3, replace $\hat{h}_t$ with the entropy of Y_t, which has 2^n possible values. However, even for modest n, the alignment encounters the small sample problem. One possible way to alleviate the difficulty is to quantize the set of all possible values of Y_t, and use entropy derived from the quantization rather than Y_t itself.

6.5 Applications of Point Processes to Complex Data. Point processes can be applied not only to spiking activity but also to other complicated data, even for continuous-valued ones. Chosen appropriately, point representations not only greatly reduce the complexity of the data, but also keep a significant amount of information. Such compact representations can be advantageous in analyzing the structure of data, such as behavior (Chi & Margoliash, 2001). In addition, as shown in this article, they can lead to efficient detection. Recent progress in computer vision and acoustics indicates that detection of complicated objects can be achieved by representing them with very simple discrete features, such as points (Amit & Geman, 1999; Amit & Murua, 2001; Amit, Koloydenko, & Niyogi, 2002; Chi, 2003; Roth, Yang, & Ahuja, 2002).

Although there are extensive possibilities of point representations for complicated data, in order for point processes to be effectively applied, several issues need to be addressed. First, what types of features should be used, and how should these features be registered? While this may not be a problem for neural activity, it poses a challenge for other types of data. Second, how should the characteristic structure be learned from sample point representations, especially when fine structure needs to be learned? This issue may have important implications for neural coding (Victor &

Purpura, 1996). Third, in the context of detection, how should the statistical significance of detected targets be assessed? As discussed earlier, this is in particular a hard problem in neural science, given that the statistical properties of neural activity in many systems are poorly understood.

Appendix

A.1 Proof of Equation 2.3. We first show that

$$L(T_x) = \sum\nolimits_{T_x\setminus(J+x)} B(t-x) \\ + \sum\nolimits_{T_x\cap(J+x)} \max_{n=1,\ldots,p} K(t-x-s_n) + C, \tag{A.1}$$

where C is the constant in equation 2.3. First, we need to find $l_a(T_x)$, the maximum likelihood of configurations that generate T_x, given T_x is generated by the template $S = \{s_1, \ldots, s_p\}$. Let X, Z_1, ..., Z_p be a configuration for T_x. By well-known results for Poisson processes (Reich, Victor, & Bruce, 1998; Brown, Barbieri, Ventura, Kass, & Frank, 2001), the likelihood of $X + x = T_x\setminus(J + x)$ is

$$Q_0 = \prod\nolimits_{t\in T_x\setminus(J+x)} f_0(t-x) \times \exp\left\{-\int_{[0,\sigma]\setminus J} f_0(\tau)\,\mathrm{d}\tau\right\}. \tag{A.2}$$

On the other hand, for each $n = 1, \ldots, p$, the likelihood of Z_n is equal to

$$Q_n = \prod\nolimits_{t\in Z_n} f_1(t) \times \exp\left\{-\int_{-\epsilon}^{\epsilon} f_1(\tau)\,\mathrm{d}\tau\right\}. \tag{A.3}$$

Because X and Z_n are independent, their joint likelihood is

$$\begin{aligned} l(X, Z_1, \ldots, Z_p) &= Q_0 \times \prod\nolimits_{n=1}^{p} \prod\nolimits_{t\in Z_n} f_1(t) \times \exp\left\{-p\int_{-\epsilon}^{\epsilon} f_1(\tau)\,\mathrm{d}\tau\right\} \\ &= Q_0 \times \prod\nolimits_{t\in T_x\cap(J+x)} f_1(t - s_{n(t)} - x) \\ &\qquad \times \exp\left\{-p\int_{-\epsilon}^{\epsilon} f_1(\tau)\,\mathrm{d}\tau\right\}, \end{aligned}$$

where $n(t)$ is the unique n such that $t \in Z_n + s_n + x$. Maximizing $l(X, Z_1, \ldots, Z_p)$ over all possible configurations that generate T_x yields

$$l_a(T_x) = P(T_x \mid T_x \text{ generated by } \{s_1, \ldots, s_n\}) = Q_0 \times Q, \tag{A.4}$$

with

$$Q := \prod\nolimits_{T_x\cap(J+x)} l(t) \times \exp\left\{-p\int_{-\epsilon}^{\epsilon} f_1(\tau)\,\mathrm{d}\tau\right\},$$

$$l(t) := \max_{n:\, t\in J_n+x} f_1(t - s_n - x).$$

Recall that $l_o(T_x) = q^n \exp\{-q\sigma\}$. Then, letting

$$C = -\int_{[0,\sigma]\setminus J} f_0(\tau)\,\mathrm{d}\tau - p\int_{-\epsilon}^{\epsilon} f_1(\tau)\,\mathrm{d}\tau + q\sigma,$$

which is independent of S,

$$\begin{aligned} L(T_x) &= \log\frac{l_a(T_x)}{l_o(T_x)} \\ &= \sum\nolimits_{t\in T_x\setminus(J+x)} \log\frac{f_0(t-x)}{q} \\ &\quad + \sum\nolimits_{t\in T_x\cap(J+x)} \max_{n:t\in J_n+x}\left\{\log\frac{f_1(t-x-s_n)}{q}\right\} + C. \end{aligned}$$

Equation A.1 is proved by combining the above equations and equation 2.1.

Now we can prove equation 2.3. By the definition of δ functions, with $T_x = T\cap[x, x+\sigma]$, it is easy to see that $R(x) = \sum_{t\in T_x} F(t-x)$. Then by equations 2.2,

$$\begin{aligned} R(x) &= \sum\nolimits_{t\in T_x\cap(J+x)} F(t-x) + \sum\nolimits_{t\in T_x\setminus(J+x)} F(t-x) \\ &= \sum\nolimits_{t\in T_x\cap(J+x)} \max_{1\le k\le p} K(t-x-s_k) + \sum\nolimits_{t\in T_x\setminus(J+x)} B(t-x). \end{aligned}$$

This combined equation A.1 then proves $L(T_x) = R(x) + C$.

A.2 Incorporation of Multiple Types of Background. Assume that a background spike sequence is a sample from one of N Poisson processes with densities $q_k(t)$. The processes may be inhomogeneous, and thus q_k may not be constant. For each k, let

$$l_k(T_x) = l(T_x|T_x \text{ is generated from } q_k),$$

still letting $l_a(T_x)$ be the maximum configuration likelihood for T_x. Then following the argument that leads to equation A.1,

$$L_k(T_x) = \log\frac{l_a(T_x)}{l_k(T_k)} = R_k(x) + C_k,$$

with $R_k(x) = \int_0^\sigma F_k(\tau)T(x-\tau)\,\mathrm{d}\tau$ and $F_k = F - \log q_k$ and F constructed from

$$K(x) = \begin{cases}\log f_1(x) & \text{if } x\in(-\epsilon,\epsilon) \\ 0 & \text{otherwise}\end{cases}, \qquad B(x) = \begin{cases}\log f_0(x) & \text{if } x\notin J \\ 0 & \text{otherwise.}\end{cases}$$

With the above changes, targets are detected at locations where all the peak responses to F_k, $k = 1, \ldots, N$ reach above-threshold local maximum.

It is also possible to develop a Bayesian approach to the detection. Let $(\pi_0, \pi_1, \ldots, \pi_n)$ be a prior probability distribution of the categories "target," "background k," $k = 1, \ldots, n$. The spike sequence T_x at a location x in the data has posteriors

$$P(T_x \text{ is a target } \mid T_x) = \frac{\pi_0 l_a(T_x)}{P(T_x)},$$

$$P(T_x \text{ is from background } k \mid T_x) = \frac{\pi_k l_k(T_x)}{P(T_x)}.$$

Then the Bayesian log-likelihood ratio of T_x being a target versus a nontarget is

$$L(T_x) = \log \frac{\pi_0 l_a(T_x)}{\sum_{k=1}^N \pi_k l_k(T_x)} = \log \frac{\pi_0}{1 - \pi_0} - \log \sum\nolimits_{k=1}^N \bar{\pi}_k \, e^{-R_k(x)},$$

with $\bar{\pi}_k = e^{-C_k} \pi_k / (1 - \pi_0)$.

A.3 Estimation of Modulated Spike Densities Based on Alignment. Let $S_1, \ldots, S_N$ be aligned templates. To estimate the density f_1 of spikes generated by template spikes, for each S_n, think of the other templates as spike sequences generated by S_n. Given $\epsilon > 0$, for each $t \in S_n$, the histogram of $|t - s|$, for $s \in S_k \cap (t - \epsilon, t + \epsilon)$, $k \neq n$, can be used to estimate the density of jittered spikes around t. Since we assume the density is the same around each template spike, we can pool all the ISIs for all $t \in S_n$ and use the histogram of

$$D_n = \bigcup\nolimits_{t \in S_n} \{|t - s| : s \in \bigcup\nolimits_{k \neq n} S_k \cap (t - \epsilon, t + \epsilon)\}$$

to estimate the density f_1. In fact, since there are $|S_n|$ time windows for the spikes in S_n, and $N - 1$ other templates, for any small interval $J \subset (-\epsilon, \epsilon)$ with duration σ,

$$\frac{1}{(N-1)\sigma |S_n|} |D_n \cap J|$$

is an estimate of the average of f_1 in J. One can pool the data of all D_n and use $D = \cup D_n$ to estimate f_1 as well.

Under the assumption of being a constant, the density f_0 of noise spikes within a target can also be estimated. Suppose the durations of the templates are all registered in $[0, \sigma]$. Then for each $n = 1, \ldots, N$,

$$\hat{f}_0^{(n)} = \frac{M_n}{(N-1)L}$$

with $L = |[0, \omega] \backslash J|$, $J = \bigcup\nolimits_{t \in S_n} (t - \epsilon, t + \epsilon)$, and

$$M_n = \sum\nolimits_{k \neq n} |\{s \in S_k : s \notin J\}|$$

gives an estimate of f_0. Moreover, $\hat{f}_0^{(n)}$ across n may be combined for estimation of f_0.

Acknowledgments

We thank Michael L. Stein for valuable comments on the manuscript. This work was supported by NIH MH59831 and MH60276.

References

Abeles, M., Bergman, H., Margalit, E., & Vaadia, E. (1993). Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *J. Neurophysiol.,70*, 1629–1638.

Abeles, M., & Gerstein, G. M. (1988). Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. *J. Neurophysiol.,60*, 909–924.

Amit, Y., & Geman, D. (1999). A computational model for visual selection. *Neural Comput., 11*(7), 1691–1715.

Amit, Y., Koloydenko, A., & Niyogi, P. (2002). *Robust acoustic object detection* (Tech. Rep. 520). Chicago: Department of Computer Science and Statistics, University of Chicago.

Amit, Y., & Murua, A. (2001). Speech recognition using randomized relational decision trees. *IEEE Trans. on Speech and Audio Processing, 9*(4), 333–341.

Baker, S. N., & Lemon, R. N. (2000). Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels. *J. Neurophysiol., 84*, 1770–1780.

Brenowitz, E. A., Margoliash, D., & Nordeen, K. W. (1997). An introduction to birdsong and the avian song system, *J. Neurobiol., 33*, 495–500.

Brown, E. N., Barbieri, R., Ventura, V., Kass, R. E., & Frank, L. M. (2001). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comput., 14*, 325–346.

Chi, Z. (2003). *Feature representation, pattern filtering, and temporal alignment for acoustic detection* (Tech. Rep. 531). Chicago: Department of Computer Science and Statistics, University of Chicago.

Chi, Z., & Margoliash, D. (2001). Temporal coding and temporal drift in brain and behavior of zebra finch song. *Neuron, 32*, 899–910.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory.* New York: Wiley.

Date, A., Bienenstock, E., & Geman, S. (1998). *On the temporal resolution of neural activity* (Tech. Rep.). Providence, RI: Division of Applied Mathematics, Brown University.

Dave, A. S., & Margoliash, D. (2000). Song replay during sleep and computational rules for sensorimotor vocal learning. *Science, 290*, 812–816.

Dave, A. S., Yu, A. C., & Margoliash, D. (1998). Behavioral state modulation of auditory activity in a vocal motor system. *Science, 282*, 2250–2253.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.

Fleuret, F., & Geman, D. (2001). Coarse-to-fine face detection. *Intl. J. Comp. Vision, 41*(1/2), 85–107.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intel., 6*(6), 721–741.

Grün, S., Diesmann, M., & Aertsen, A. (2002a). Unitary events in multiple single-neuron spiking activity: I. Detection and significance stimulus properties. *Neural Comput., 14*(1), 43–80.

Grün, S., Diesmann, M., & Aertsen, A. (2002b). Unitary events in multiple single-neuron spiking activity: II. Nonstationary data. *Neural Comput., 14*(1), 81–120.

Hahnloser, R. H. R., Kozhevnikov, A. A., & Fee, M. S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature, 419*, 65–70.

Louie, K., & Wilson, M. A. (2001). Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron, 29*, 145–156.

Margoliash, D. (2001). Do sleeping birds sing? Population coding and learning in the bird song system. *Prog. Brain Res., 130*, 319–331.

Margoliash, D. (2002). Evaluating theories of bird song learning: Implications for future directions. *J. Comp. Physiol. A, 188*, 851–866.

Mooney, R. (2000). Different subthreshold mechanisms underlie song selectivity in identified HVc neurons of the zebra finch. *J. Neurosci., 20*(14), 5420–5436.

Nádasdy, Z., Hirase, H., Czurkó, A., Csicsvari, J., & Buzsáki, G. (1999). Replay and time compression of recurring spike sequences in the hippocampus. *J. Neurosci., 19*(21), 9497–9507.

Reich, D. S., Victor, J. D., & Bruce, W. K. (1998). The power ratio and the interval map: spiking models and extracellular recordings. *J. Neurosci., 18*(23), 10090–10104.

Riehle, A., Grün, S., Diesmann, M., & Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science, 278*, 1950–1953.

Rieke, F., Warland, D., de Ruyter van Steveninck, R. R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press.

Rojer, A. S., & Schwartz, E. L. (1992). A quotient space Hough transform for space-variant visual attention. In G. A. Carpenter & S. Grossbert (Eds.), *Neural networks for vision and image processing*. Cambridge, MA: MIT Press.

Roth, D., Yang, M.-H., & Ahuja, N. (2002). Learning to recognize three-dimensional objects. *Neural Comput., 14*(5), 1071–1103.

Shea, S. D., Rauske, P. L., & Margoliash, D. (2001). Identification of HVc projection neurons in extracellular records by antidromic stimulation. *Soc. Neurosci. Abstr., 27*, 381–386.

Skaggs, W. E., & McNaughton, B. L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science, 271*, 1870–1873.

Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., & Aertsen, A. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature, 373*, 515–518.

Victor, J. D., & Purpura, K. P. (1996). Nature and precision of temporal coding in visual cortex: A metric-space analysis. *J. Neurophysiol., 76*(2), 1310–1326.

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science, 265*, 676–679.

Yu, A. C., & Margoliash, D. (1996). Temporal hierarchical control of singing in birds. *Science, 273*, 1871–1875.

Received October 11, 2002; accepted April 3, 2003.