

Cortical Discrimination of Complex Natural Stimuli: Can Single Neurons Match Behavior?

Le Wang,^{1,2} Rajiv Narayan,^{1,2} Gilberto Graña,^{1,2} Maoz Shamir,^{1,2,3} and Kamal Sen^{1,2}

¹Hearing Research Center, Department of Biomedical Engineering, ²Center for Biodynamics and Program in Mathematical and Computational Neuroscience, and ³Department of Mathematics, Boston University, Boston, Massachusetts 02215

A central finding in many cortical areas is that single neurons can match behavioral performance in the discrimination of sensory stimuli. However, whether this is true for natural behaviors involving complex natural stimuli remains unknown. Here we use the model system of songbirds to address this problem. Specifically, we investigate whether neurons in field L, the homolog of primary auditory cortex, can match behavioral performance in the discrimination of conspecific songs. We use a classification framework based on the (dis)similarity between single spike trains to quantify neural discrimination. We use this framework to investigate the discriminability of single spike trains in field L in response to conspecific songs, testing different candidate neural codes underlying discrimination. We find that performance based on spike timing is significantly higher than performance based on spike rate and interspike intervals. We then assess the impact of temporal correlations in spike trains on discrimination. In contrast to widely discussed effects of correlations in limiting the accuracy of a population code, temporal correlations appear to improve the performance of single neurons in the majority of cases. Finally, we compare neural performance with behavioral performance. We find a diverse range of performance levels in field L, with neural performance matching behavioral accuracy only for the best neurons using a spike-timing-based code.

Key words: auditory cortex; speech; birdsong; discrimination; behavior; field L

Introduction

The analysis of the performance of single cortical neurons on perceptual tasks has played a central role in understanding the link between the brain and behavior (for review, see Parker and Newsome, 1998; Romo and Salinas, 2003). This is appropriate because a clear understanding of the capacity of single neurons is a prerequisite for understanding any neural code, be it a single neuron code or a population code. A striking and well established result in many cortical brain areas is that the performance of single cortical neurons can match the behavioral performance of an animal (Britten et al., 1992; Hernandez et al., 2000). Previous studies used relatively simple synthetic stimuli to probe both neural and behavioral performance. However, cortical neurons can display highly nonlinear responses when probed using complex natural stimuli (Theunissen et al., 2000; Bar-Yosef et al., 2002; David et al., 2004; Machens et al., 2004; Felsen et al., 2005; Sharpee et al., 2006). Thus, whether single cortical neurons can match behavioral performance in a task involving complex natural stimuli remains unknown.

The combined knowledge of behaviorally relevant stimuli and the underlying neural circuitry make songbirds an attractive sys-

tem with particular relevance for human speech (Doupe and Kuhl, 1999). In particular, field L, the avian homolog of primary auditory cortex, provides a model for investigating the cortical processing of natural stimuli, e.g., birdsongs. As a population, field L neurons show selectivity for conspecific songs (Grace et al., 2003). Modulation tuning of neural ensembles in field L facilitates discrimination across different classes of sounds as well as within the class of conspecific sounds (Woolley et al., 2005). Thus, field L provides an ideal test bed for comparing neural and behavioral discrimination of a behaviorally important class of sounds, i.e., conspecific songs. However, neural discrimination performance at the single neuron level in field L remains poorly understood.

Behaviorally, songbirds can discriminate accurately between songs based on a single presentation (Cynx, 1993; Shinn-Cunningham et al., 2006). In this situation, the available sensory information consists of a single spike train in response to a song from each neuron in the relevant population. Thus, to assess the contribution of single neurons to behavior, it is important to quantify the information available from single spike trains. Previous studies have revealed that the information in single spike trains is present at fine timescales (Wright et al., 2002), and the optimal timescale for neural discrimination of conspecific songs is ~10 ms (Narayan et al., 2006). These observations suggest that a spike-timing-based code could mediate song discrimination. However, several questions remain open. How does the accuracy of discrimination based on a single spike train using a spike-timing code compare with other candidate codes, i.e., rate and interval codes? How does temporal correlation impact performance based on a spike-timing code? How does average neural

Received Aug. 24, 2006; revised Nov. 16, 2006; accepted Dec. 4, 2006.

This work was supported by National Institute on Deafness and Other Communication Disorders Grant 1R01 DC-007610-01A1. We thank Steve Colburn, Conor Houghton, and Haim Sompolinsky for discussions and Sabina Khan for assistance with data analysis.

Correspondence should be addressed to Kamal Sen, Hearing Research Center, Department of Biomedical Engineering, Center for Biodynamics, Program in Mathematical and Computational Neuroscience, Boston University, 44 Cummings Street, RM 414B, Boston, MA 02215. E-mail: kamalsen@bu.edu.

DOI:10.1523/JNEUROSCI.3699-06.2007

Copyright © 2007 Society for Neuroscience 0270-6474/07/270582-08\$15.00/0

performance compare with the performance of the best neurons? Which codes are consistent with behavioral performance? Here we address these questions, using extensive computational analyses of experimental data and modeling.

Materials and Methods

The experimental methods have been described in detail previously (Sen et al., 2001; Narayan et al., 2006). Here we give an overview of the dataset and focus on describing the computational methods and modeling.

Neural data

The data analyzed here come from a previous study in which we recorded responses from field L of anesthetized and awake adult male zebra finches (Narayan et al., 2006). For each recording site, we obtained 10 trials of neural responses to 20 conspecific zebra finch songs. The data were categorized into three groups based on the type of recording and isolation of the spike waveforms: anesthetized single unit ($n = 6$), anesthetized multiunit ($n = 18$), and awake multiunit ($n = 14$). As reported in the previous study, the multiunits comprised small clusters of neurons (approximately two to five) dominated by single units. There were no statistically significant differences in discrimination performance, optimal temporal resolution, and temporal integration time constant between the three groups (Narayan et al., 2006).

Computational methods

Potential alternatives for quantifying neural discrimination. To quantify neural performance, we adopted a metric-based spike train classification framework, although we did consider alternative methods. The same problem could be addressed, in principle, using ideal observer analysis (Green and Swets, 1966) or information theory (Rieke et al., 1997). In this study, we were interested not only in the information present in spike counts or average rate but also at fine temporal resolutions. Ideal observer analysis can be used to obtain discriminability measures on a fine timescale. Specifically, one can compute the log likelihood ratio for two different stimuli based on observed spike counts in a small time bin. If spike counts are statistically independent across different time bins, then the total discriminability accumulated over time can be obtained simply by summing the binwise log likelihood ratios. However, the statistical independence assumption may not be satisfied by experimental data. Indeed, the spike trains in our dataset contained significant temporal correlations (see Results), which complicate the application and interpretation of ideal observer analysis. Nevertheless, we compared the performance of the classifier with the ideal observer in a computational model (supplemental data, available at www.jneurosci.org as supplemental material).

Another potential approach is information theory, a powerful model-independent framework for quantifying neural performance. Information present at fine temporal resolutions can be estimated using the so-called “direct” method for computing mutual information. This method requires estimating the probability distributions of different spike sequences or “words” in the neural response (Strong et al., 1998; Wright et al., 2002). Our dataset, which consisted of 10 trials for each song, was insufficient for estimating the word distributions necessary for applying this method, especially for the long words necessary to probe the relatively long integration timescale on the order of hundreds of milliseconds (Narayan et al., 2006).

Having considered these alternatives for quantifying neural performance, we adopted a spike train classifier based on (dis)similarity measures for single spike trains (Victor and Purpura, 1997; Machens et al., 2003; Narayan et al., 2005, 2006). This approach makes no assumptions about the statistical structure of the spike trains, e.g., Poisson statistics. It is applicable to single spike trains and does not require estimating the underlying probability distribution of spike trains. It is flexible and can be used in conjunction with different (dis)similarity measures between spike trains. Data from many cortical neurophysiology laboratories have similar constraints, i.e., presence of temporal correlations and limited numbers of trials. Spike train classification provides a computational framework for the analysis of such datasets.

Spike train classification. In a previous study, we used a classification

method based on the van Rossum spike distance metric (VR) to quantify timescales underlying discrimination (van Rossum, 2001; Machens et al., 2003; Narayan et al., 2006). In that approach, a template spike train was chosen from one of the 10 trials for each song. The remaining spike trains were assigned to the song with the closest template based on VR. This procedure was repeated 100 times for different templates. The percentage of correctly classified spike trains (percentage correct) was used as a measure of discrimination. The chance level for classification was 5% because a spike train could be assigned to 1 of 20 songs.

For this study, we expanded the classification framework by exploring VR further and by investigating other (dis)similarity measures, i.e., the Victor and Purpura spike timing metric (VP_{spike}), the Victor and Purpura spike interval metric (VP_{interval}), and a correlation-based similarity measure, R_{corr} (Schreiber et al., 2003).

VR metric. The VR metric quantifies the dissimilarity between pairs of spike trains by first filtering them using a decaying exponential kernel with a time constant τ :

$$f(t) = \sum_{i=1}^M H(t - t_i) e^{-\frac{(t-t_i)}{\tau}}, \quad (1)$$

where t_i is the i th spike time, M is the total number of spikes, and $H(t)$ is the Heaviside step function. The spike distance is then computed as the Euclidean distance (integral of the squared difference) between a pair of filtered spike trains, f and g :

$$D^2(f, g)_\tau = \frac{1}{\tau} \int_0^\infty [f(t) - g(t)]^2 dt. \quad (2)$$

The parameter τ can be varied to examine discrimination over different timescales of the neural response (van Rossum, 2001).

Victor–Purpura metrics (VP_{spike} and VP_{interval}). Victor and Purpura (1997) introduced a family of spike distance metrics that measure dissimilarity between two spike trains in terms of the minimum cost of transforming a spike train into another spike train through a series of elementary operations. For all of the metrics, two permitted operations are the addition of a single spike and the deletion of a single spike, both for a cost of 1. The metric based on spike times (VP_{spike}) allows a spike to be shifted by an amount dt for a cost of qdt . The parameter q has units of time^{-1} , and the quantity $1/q$ is a measure of the temporal resolution of the metric. The metric based on spike intervals (VP_{interval}) permits changing of the length of an interspike interval by an amount dt for a cost qdt . An important difference between the two metrics is that the shifting of an interspike interval in VP_{interval} changes the spike times of all subsequent spikes, whereas the shifting of a spike time in VP_{spike} only causes a change in the intervals immediately preceding and after the shifted spike. Additionally, (dis)similarities based on VP_{spike} are not necessarily equivalent to (dis)similarities based on VP_{interval} . For example, random deletions of spikes in the data tend to reduce performance based on VP_{interval} more than VP_{spike} (Victor and Purpura, 1996).

Correlation-based similarity measure (R_{corr}). The R_{corr} measure was based on a recently proposed correlation-based measure of spike similarity (Schreiber et al., 2003). The similarity between two spike trains, $\vec{\tau}_i$ and $\vec{\tau}_j$, was calculated as follows:

$$R_{\text{corr}} = \frac{\vec{\tau}_i \cdot \vec{\tau}_j}{|\vec{\tau}_i| |\vec{\tau}_j|}, \quad (3)$$

where $\vec{\tau}_i$ and $\vec{\tau}_j$ were obtained by filtering $\vec{\tau}_i$ and $\vec{\tau}_j$ using a Gaussian filter with mean 0 and SD σ . A value close to 1 indicates similar spike trains, whereas a value close to 0 is indicative of dissimilarity. The width of the filter was adjusted so that $2.8\sigma \approx \tau$ of VR (see below, Analysis of kernel shape in VR).

Performance, biological interpretability, and choice of metrics. We compared the classification performance attained with the different (dis)sim-

ilarity measures. Mean \pm SEM performances ($n = 38$) for all of the recorded sites were 48.5 ± 3.8 , 44.6 ± 3.8 , and $60.4 \pm 4.1\%$ for VR, VP_{spike} , and R_{corr} , respectively. Of the three measures, VR was the easiest to interpret biologically. In essence, the overall computation of VR consists of three steps: filtering, comparison, and integration (see Eqs. 1, 2). Biophysically, the filtering could be accomplished at a synapse. Interestingly, the optimal timescale for the filter is typically ~ 10 ms for the data, which is well matched to synaptic timescales. Several biophysical mechanisms have also been described for temporal integration (for review, see Major and Tank, 2004). Although the precise mechanisms of the comparison step remains unclear, it seems plausible that a comparison occurs during song identification. R_{corr} is similar to VR mathematically (one can think of the two measures as the cosine of the angle and the distance between two vectors, respectively). Although R_{corr} outperformed VR, the difference in performance was attributable to the normalization factor in R_{corr} (data not shown), which is difficult to interpret biologically. The VP_{spike} measure was also relatively difficult to interpret in terms of biological mechanisms. Based on these considerations, we used a classifier based on VR to evaluate neural performance based on spike timing and rate. The spike timing measure (VR_{timing}) was defined as VR at a time resolution of 10 ms, and the spike rate measure (VR_{rate}) was VR at a resolution of 1000 ms. To evaluate interval-based coding, we used a classifier based on VP_{interval} , the only such measure in the literature.

Analysis of kernel shape in VR. We examined how the shape of the filter kernel influenced the performance of the VR metric. Three different kernels were used: a decaying exponential with a time constant τ , a Gaussian with an SD σ , and an alpha function with parameter α . The areas of all kernels were normalized to 1. To obtain equivalent timescales with the different kernels, the full width of each kernel at a height of $1/e$ was adjusted to be equal to τ . This yielded the relationship $\tau \approx 2.8\sigma \approx 3.0/\alpha$.

Analysis of effects of jitter in onset. The effect of onset jitter on the classification performance was studied by introducing a fixed time shift to every spike time in each trial. The time shifts for each trial were independently sampled from a 0 mean Gaussian distribution whose SD was varied from 0 to 20 ms.

In addition to examining the effects of random onset jitter on performance, we searched for the “optimal” onset alignment of the spike trains that would yield the best performance. For each stimulus, the first trial spike train was selected as a reference, and the best relative shift of each of the remaining nine trials was determined using a grid search algorithm, to maximize the value of R_{corr} averaged over all pairs of spike trains (Schreiber et al., 2003). The R_{corr} values were computed from 2-s-long spike trains (beginning at the stimulus onset) that were convolved with an exponential kernel with a decay time constant of 10 ms. Because this optimization problem was combinatorially explosive even for small numbers of trials, the range of possible shifts was restricted to $[-10, 10]$ ms with a step size of 2 ms. This allowed for a maximal relative shift of 20 ms between a pair of spike trains. The grid search results were further refined using an iterative grid-walk procedure to find the nearest local maximum. During each iteration, a shift of ± 1 ms was considered along each of the nine possible search dimensions, and the optimal shift was updated based on the dimension that would produce the maximum improvement in the mean R_{corr} . The procedure was repeated until there was no improvement in R_{corr} . The SD of the optimal shifts yields an estimate of the onset jitter.

Analysis of effects of temporal correlations. To reduce “within-trial” temporal correlations, spike trains were binned at 1 ms intervals, and the bins were randomly shuffled across trials. Shuffling typically reduced negative correlations at small lags and increased the number of short-duration interspike intervals (data not shown). The discrimination performance of the shuffled data were compared with the performance of the original data using the VR metric.

We compared the autocorrelation histogram before and after shuffling to relate the effects of shuffling on temporal correlations to changes in discrimination performance. The autocorrelation histogram for each unit was computed using standard methods (Dayan and Abbott, 2001) and was normalized to unity at zero lag. We then used a jackknife resampling method to correct for bias and compute the SD of the estimate.

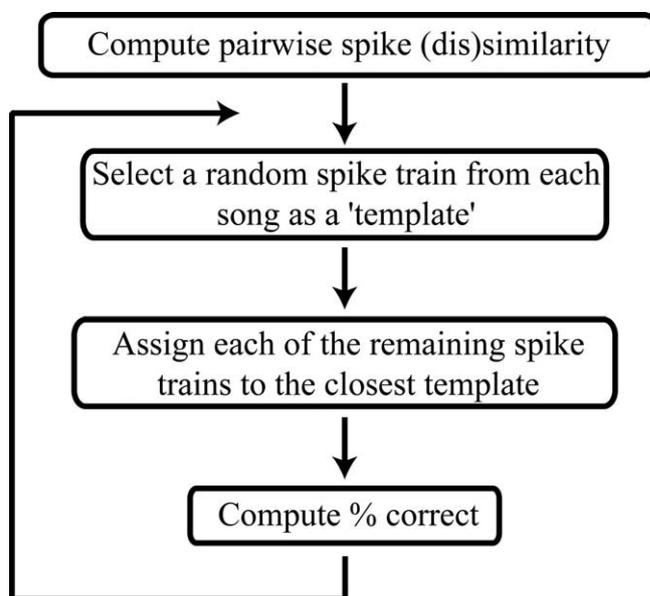


Figure 1. Song classification based on neural responses. A schematic of the supervised classification procedure used to obtain a measure of neuronal performance using a spike distance metric.

Typically, both bias and SD of the estimates were small.

We explored several measures to quantify the differences in the autocorrelation before and after shuffling, e.g., difference taken at the negative peak of the autocorrelation from the real data, as well as the difference in area between the two autocorrelations both with and without considering the signs for the area for a wide range of lags. Here we report the measure that was most strongly correlated with performance: the signed difference in area between the two autocorrelations from 0 to 10 ms lags (D_{area}). The correlation between D_{area} and the change in discrimination performance after shuffling was computed across all of the units. One unit with a very low firing rate did not show the characteristic dip in the autocorrelation function (see Fig. 5A) but showed an unusual offset in the baseline values of the real and shuffled autocorrelations, which would have produced an artifact in the D_{area} measure. For this unit, D_{area} was set to 0.

Behavioral data

The behavioral measures of performance are based on Cynx (1993) and Shinn-Cunningham et al. (2006). These studies reported the behavioral accuracy of zebra finches in discriminating between conspecific songs, finding near perfect performance levels in the range of 90–100%. In these studies, the chance level of performance for zebra finches was 50%. To compare the behavioral range with the neural performance levels in this study in which the chance level was 5%, we plotted both the neural data and behavioral range of performances between chance level and perfect performance, a method used in psychoacoustics (Shinn-Cunningham et al., 2006).

Results

A major goal of this study was to quantify neural discrimination performance in field L in the discrimination of conspecific songs. To address this problem, we used a classification framework for single spike trains in response to songs based on a spike distance metric (Fig. 1) (see Materials and Methods). Before proceeding to quantify neural performance, we investigated several additional aspects of the spike train classifier that were not addressed in previous studies (supplemental data, available at www.jneurosci.org as supplemental material).

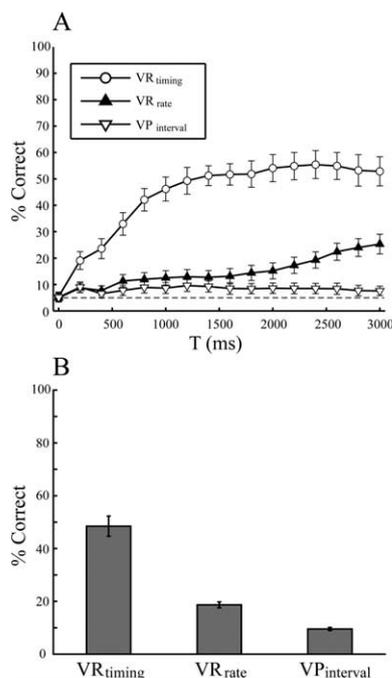


Figure 2. *A*, A comparison of accuracy as a function of spike train length for a single site, when the discrimination is based on spike timing (VR_{timing} is VR at a resolution of 10 ms), spike rate (VR_{rate} is VR at a resolution of 1000 ms), or interspike intervals (VP_{interval} at 10 ms resolution). The error bars represent 1 SD computed from 100 repetitions of the song classification procedure (Fig. 1). Chance level for classification is 5% (dashed line). *B*, Mean performance accuracy (± 1 SEM; $n = 38$) for all recording sites using the three different codes. The performance of the three codes were significantly different from each other ($T = 2$; $p < 0.05$, repeated-measures ANOVA followed by Scheffé's multiple comparison test).

Discrimination based on spike timing, spike rate, and interspike intervals

We examined the differences in accuracy when songs were classified using information present in spike timing, rate, or interspike intervals (see Materials and Methods). Figure 2*A* shows the performance as a function of spike train duration for a representative site in our data. The mean accuracies (± 1 SEM; $n = 38$) achieved were 48.4 ± 3.8 , 18.7 ± 1.1 , and $9.5 \pm 0.6\%$ for spike timing, rate, and spike intervals, respectively (Fig. 2*B*). Thus, classification performance based on spike timing was better compared with rate and interspike interval. Here it may be useful to remind the reader that we have used the term spike-timing code specifically to refer to the presence of information at short time-scales (~ 10 ms) relative to the rate code (1000 ms).

Robustness and constraints for biological implementation

The classification scheme makes some assumptions that may not be satisfied exactly in a biological system. One of these assumptions is that the spike trains being compared are aligned at the onset of stimuli. We next explored how the performance changed as this ideal assumption was relaxed (for the similar analysis for other assumptions, see supplemental data, available at www.jneurosci.org as supplemental material).

We tested the performance of the classifier at different levels of onset jitter. Figure 3*A* shows that the discrimination accuracy was $48.4 \pm 3.8\%$ (mean \pm SEM; $n = 38$) with no onset jitter, and it decreased gradually as the level of onset jitter was increased to 20 ms, when the accuracy dropped to $22.2 \pm 1.7\%$. Thus, performance degraded gracefully, with a statistically significant decrease in performance noticeable only above an onset jitter of 5

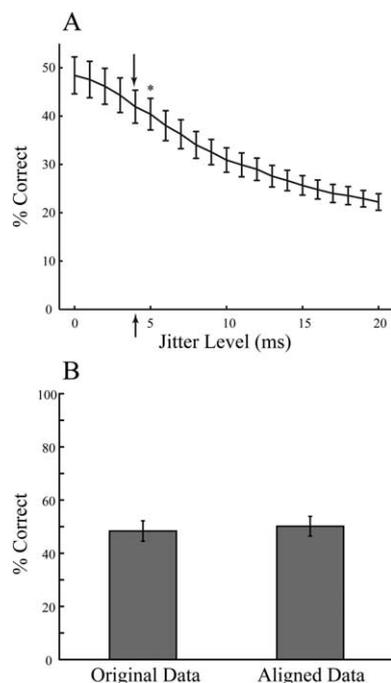


Figure 3. Onset jitter. *A*, This figure illustrates the effect of onset jitter on the neural performance (mean \pm SEM; $n = 38$). The jitter level represents the SD of the Gaussian distribution that was randomly sampled to obtain the amount of jitter that was added to each trial in the dataset. The performance was measured using VR ($T = 2$ s; $\tau = 10$ ms). A statistically significant decrease in performance was observed at an onset jitter of 5 ms or greater ($*p < 0.05$, repeated-measures ANOVA followed by Scheffé's multiple comparison test). The onset jitter in field L responses, estimated using a de-jittering procedure (see Materials and Methods), was 3.9 ms (arrows). *B*, A comparison of discrimination accuracy (mean \pm SEM; $n = 38$) obtained using the original spike trains and the de-jittered spike trains (see Materials and Methods). The difference was statistically significant ($p < 0.05$, paired t test).

ms when the accuracy was $40.4 \pm 3.3\%$. This provides a constraint on the accuracy of alignment necessary in a biological implementation to match the performance of the classifier. In addition to examining the effect of adding onset jitter, we also investigated the effect of “de-jittering” the spike trains by shifting each trial to maximize the average pairwise cross-correlation between trials (see Materials and Methods). After de-jittering, performance improved to $50.2 \pm 3.7\%$ (Fig. 3*B*). In the auditory system, onset cues are precise, highly salient, and represented at multiple levels (for review, see Phillips et al., 2002). Such cues may facilitate the temporal alignment of a sensory input signal with a stored template signal during a classification process. In our dataset, the onset jitter of responses was estimated to be 3.9 ms, using the de-jittering procedure (Fig. 3*A*) (see Materials and Methods).

Effects of temporal correlations

We tested the effect of temporal correlations in the spike trains by shuffling the spike trains across trials (see Materials and Methods). The discrimination accuracy curve for a representative neuron is shown in Figure 4*A*. The peak accuracy was 55.6% before shuffling, and it reduced to 40.3% after shuffling. Thus, for this site, temporal correlations present in the original spike trains enhanced discrimination performance. The majority of sites (23 of 38) showed an increase in performance in the presence of temporal correlations (average increase of 10.7%), whereas the remaining sites showed a decrease (average decrease of 4.2%) (Fig. 4*B*).

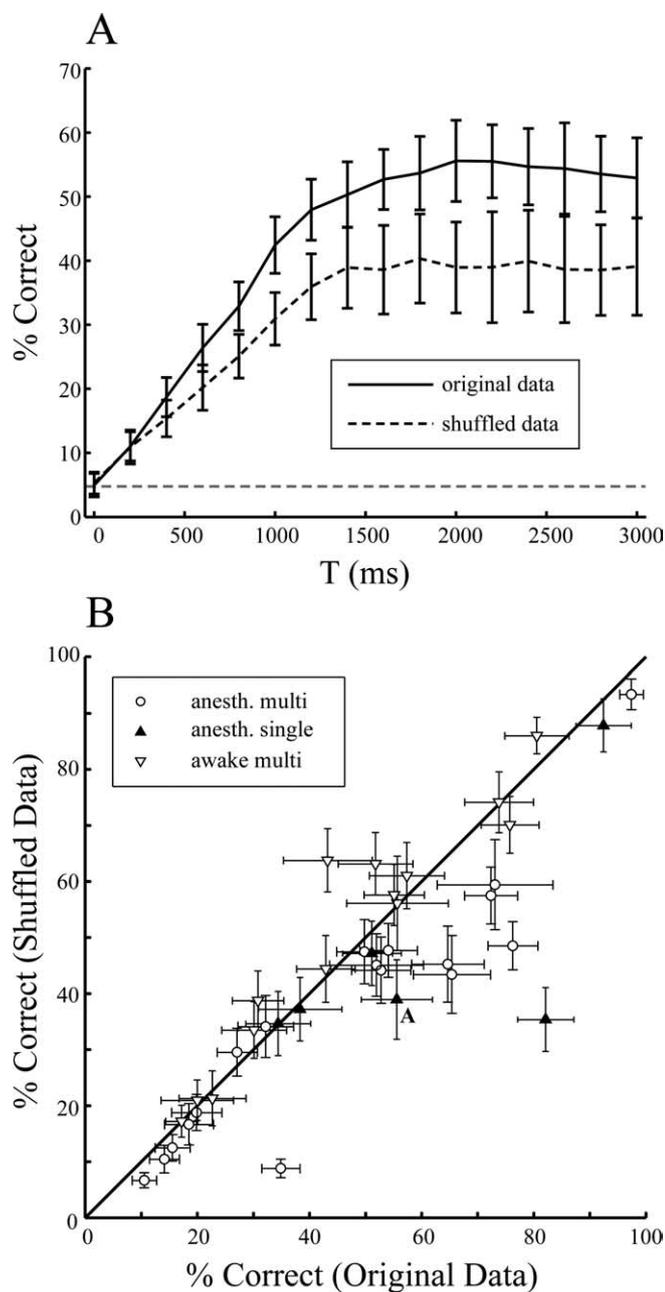


Figure 4. *A*, A comparison of discrimination accuracy at a single site for real data (solid line) and shuffled data (dashed line; see Materials and Methods). *B*, The mean accuracy obtained from the shuffled spike trains versus the performance of the unshuffled spike trains for all sites (± 1 SD; $n = 38$). The site shown in *A* is indicated (A). The performance was measured using VR ($T = 2$ s; $\tau = 10$ ms). The performance based on the shuffled data were significantly different from the performance based on the original data ($T = 2$ s; $p < 0.05$, paired *t* test).

The autocorrelation histogram of the original spike trains typically showed a negative dip for lags between 0 and 10 ms relative to the autocorrelations of the shuffled data (Fig. 5*A*). At longer lags, the two histograms were similar to each other. The difference in the area of the autocorrelations, D_{area} (see Materials and Methods) (Fig. 5*A*, inset), was negatively correlated with the change in discrimination accuracy before and after shuffling and had a correlation coefficient of -0.81 (Fig. 5*B*).

Neural versus behavioral performance

Figure 6 illustrates that the classification accuracy for the 38 neural sites we recorded from ranged from 10.5 to 97.4%, with a

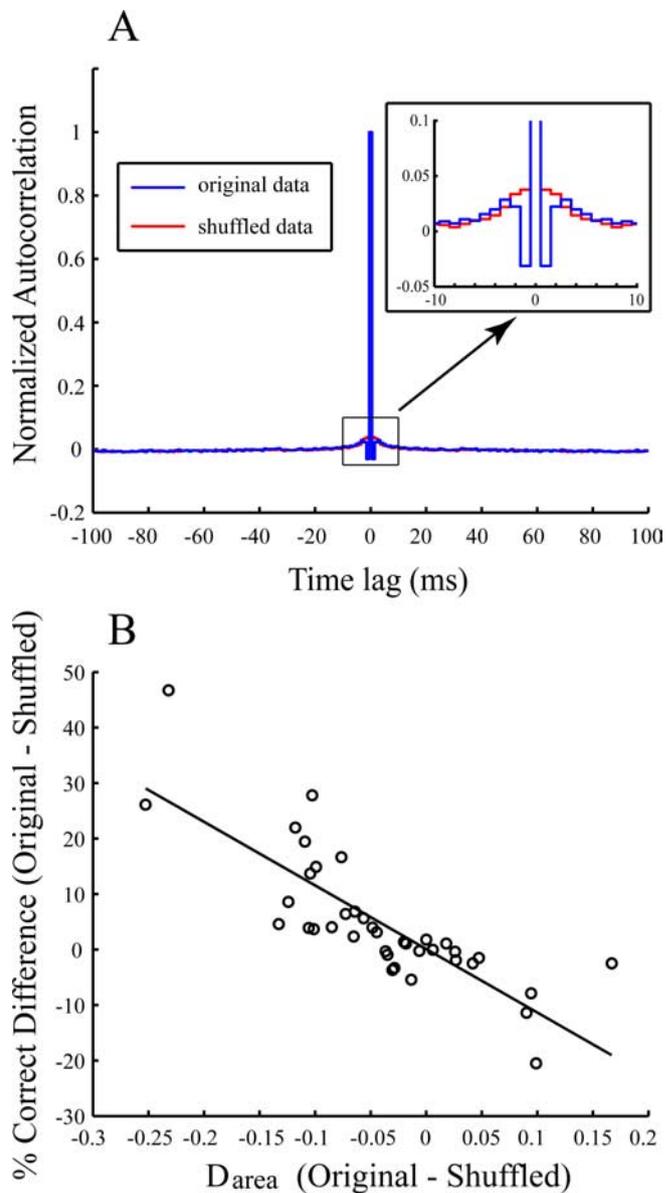


Figure 5. *A*, The autocorrelation histograms for real data (blue) and shuffled data (red) from a representative unit plotted for lags ranging from -100 to 100 ms (bin size of 1 ms). The inset is a zoomed-in version of the autocorrelation histograms, showing the region from -10 to 10 ms. Note the negative dip in the autocorrelation in the range of lags between 1 and 10 ms (inset). *B*, The relationship between the difference in discrimination accuracy before and after shuffling and D_{area} (see Materials and Methods). Each circle represents one unit. The correlation was -0.81 ($p < 0.01$; $n = 38$). The least-squares linear regression fit to the data is also shown.

mean of 48.5% (VR, $T = 2$ s; $\tau = 10$ ms). The classification performance for the three candidate codes we considered i.e., spike timing, spike rate, and interspike intervals, are shown for both the average and the best neuron (see Discussion). The range of behavioral accuracy reported for zebra finches performing a discrimination task on conspecific songs is also indicated (gray shaded region). Neural performance was within the range of behavioral performance levels only for the best neurons using spike-timing information.

Discussion

Can single neurons match behavioral discrimination of natural stimuli?

We considered several candidate neural codes, i.e., spike timing, rate, and interval, for song discrimination. Performance based on

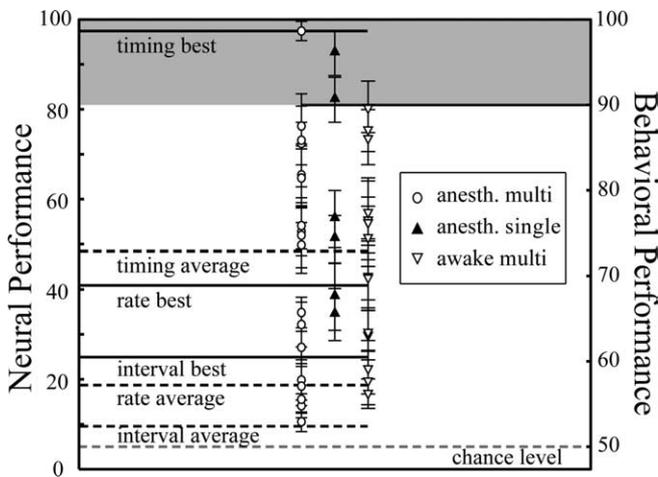


Figure 6. The figure shows the mean neural accuracy (± 1 SD; $n = 38$) for individual sites of anesthetized multiunit, anesthetized single unit (open circles and filled triangles), and awake multiunit data (open triangles) computed using the VR distance ($T = 2$ s; $\tau = 10$ ms). The best (solid horizontal lines) and mean (dashed horizontal lines) neural performance based on spike timing, rate, and interspike intervals are indicated against the left vertical scale. The range of behavioral discrimination performance is indicated by the gray region against the right vertical scale (Cynx, 1993). The chance level for classification is 5% for the neural data and 50% for the behavioral data (gray dashed line).

a spike-timing code was highly diverse, spanning a range from near-chance level to near perfect. Part of the reason for such a broad range of performances, including neurons that performed poorly, may be that we did not “tune” the stimuli for each neuron separately but sampled all neurons using the same stimuli (Purushothaman and Bradley, 2005). Such a sampling strategy is more reflective of the situation during natural behavior in which an incoming stimulus impinges on a population of neurons with a broad range of stimulus selectivities, including those that may not be optimally tuned for processing the given stimulus. On average, neural discrimination based on a spike-timing code outperformed neural discrimination based on rate and interval codes. This is consistent with growing evidence that single auditory cortical spike trains can contain significant amounts of information at a fine temporal resolution (Furukawa and Middlebrooks, 2002; Wright et al., 2002; Lu and Wang, 2004; Nelken et al., 2005; Narayan et al., 2006; Schnupp et al., 2006). Although spike-timing-based codes can be potentially susceptible to onset jitter, we found that the average onset jitter in the neural responses in our dataset was smaller than jitter levels that produced a significantly detectable drop in discrimination performance (Fig. 3A).

In some previous studies, behavioral performance has been found to be more strongly correlated with the performance of the best neurons rather than the average neuron, as posited by the lower envelope principle (Parker and Newsome, 1998). Thus, for each of the candidate codes, we also examined the performance of the best neurons in our dataset. We found that neural performance could match behavioral performance of songbirds in a song discrimination task only for the best neurons using a spike-timing code. This finding can be contrasted with previous findings in motion direction discrimination in middle temporal area MT (Parker and Newsome, 1998) and flutter vibration discrimination in the somatosensory cortex (Romo and Salinas, 2003). In both cases, a rate code was sufficient to explain behavioral performance. Our finding is consistent with the spirit of the lower envelope principle, extending its application beyond rate coding

(i.e., based on long timescales on the order of seconds) of synthetic stimuli to spike-timing-based coding (i.e., based on shorter timescales on the order of 10 ms) of natural stimuli.

Effect of temporal correlations on neural performance

Although the role of correlations in a population code has received much attention, the role of temporal correlations on the performance of single neurons is less well understood. In the context of population coding, correlations have generally been thought to limit cortical performance (Zohary et al., 1994), although this need not be the case (Romo et al., 2003). In the temporal domain, an elegant study in area MT by Osborne et al. (2004) demonstrated that the performance of single neurons in direction discrimination was reduced by temporal correlations in the spike train (Osborne et al., 2004). In this study, we used an approach inspired by Osborne et al. (2004) to compare the performance based on the real spike trains versus shuffled spike trains, which did not contain within-trial temporal correlations (see Materials and Methods). This analysis revealed that the temporal correlations present in single spike trains improved neural performance in the majority of cases. This is not directly comparable, nor contradictory, to the results of Osborne et al. (2004) because of many differences between the two studies. In particular, Osborne et al. (2004) used stimuli with constant amplitude, i.e., speed, whereas we used natural stimuli in which the amplitude, i.e., amplitude envelope of songs, varied with time. A second difference is that Osborne et al. (2004) examined the effects of temporal correlations on spike counts in increasingly longer windows, whereas we considered a spike-timing-based code.

An analysis of the spike train autocorrelations revealed a specific feature that was significantly correlated with the enhancement in performance: the signed area of the negative dip in the autocorrelation histogram within a range of lags between 0 and 10 ms (Fig. 5). Larger negative areas were correlated with better performance. A dip in the autocorrelation arises, fundamentally, as a result of a decrease in the probability of spiking at short lags (relative to the autocorrelation of the shuffled data). Our results indicate that such a decrease in the spike train autocorrelation can increase discrimination accuracy of a spike-timing-based code. Interestingly, studies on cortical coding that have found correlations to limit performance have mainly considered the effect of positive correlations on a rate code (Zohary et al., 1994; Osborne et al., 2004). Thus, the effects of correlation on coding accuracy can depend on the particular form of the correlation as well as the specific type of code being considered (Abbott and Dayan, 1999; Sompolinsky et al., 2001). A recent theoretical study investigated the information content of the dynamic response of single cells to natural stimuli (Shamir et al., 2007). The study found that positive temporal correlations reduced the information content of the neural response, whereas negative temporal correlations increased the information content relative to the uncorrelated (shuffled) case. This basic intuition is consistent with the results of Osborne et al. (2004), as well as our current findings. A plausible candidate mechanism underlying the dip in the autocorrelation is refractoriness (Berry and Meister, 1998; Schaette et al., 2005), although fast delayed inhibition (Wehr and Zador, 2003; Narayan et al., 2005) may also contribute. Previous studies in the sensory periphery using time-varying stimuli have demonstrated that temporal correlations can decrease the trial-to-trial variability of neural responses (Berry and Meister, 1998; Schaette et al., 2005) and improve performance (Ratnam and Nelson, 2000; Chacron et al., 2001). Our study provides examples at the

cortical level, in which correlations can improve rather than degrade neural performance.

Recognition and readout

The performance of the spike train classifier can be thought of as an estimate of the information available at the level of field L for song identification. Our analysis suggests that information present at a fine timescale is important for such identification. This information may be extracted by downstream circuitry and represented in terms of the firing rates of single neurons to create a simple readout for recognition. In such a scheme, a downstream “readout” neuron would fire only when a particular song was presented, representing the output of the classification process. Neurons in areas downstream from field L, e.g., HVC (used as a proper name), interfacial nucleus of the nidopallium (NIf), and caudal mesopallium (cM), which show selectivity for specific songs, may represent such readout neurons. However, the biological mechanisms underlying the readout remain unclear. Here, theoretical work on the readout of spike-timing codes (Buonomano, 2000; Hopfield and Brody, 2001; Hopfield, 2004; Gutig and Sompolinsky, 2006) can guide the formulation of experimental hypotheses. A promising experimental approach for testing such hypotheses may be to record simultaneously in the input and the readout areas, with intracellular recordings in the readout area, as illustrated elegantly in the olfactory system (MacLeod et al., 1998; Perez-Orive et al., 2002) and in a recent study in the birdsong system (Coleman and Mooney, 2004). Similar recordings in field L and downstream areas, e.g., cM, NIf, and HVC, may provide new insights into biological mechanisms underlying the readout of spike-timing codes.

Limitations and future directions

Our study took only the first steps toward understanding the relationship between neural responses and behavior for natural stimuli. One of the limitations of this study is that the neural and behavioral performance measures were obtained from separate studies (Cynx, 1993; Best et al., 2005; Narayan et al., 2006; Shinn-Cunningham et al., 2006). Ideally, this comparison should be performed simultaneously in the same subject. In addition to a more careful comparison between neural and behavioral performance using the same stimuli in the same subject on a trial-by-trial basis, such simultaneous experiments allow the exciting possibility of microstimulating neurons to bias perception (for review, see Cohen and Newsome, 2004). Such experiments may establish a firmer link between neural and behavioral performance. Although we focused on single neuron codes, information about auditory stimuli is typically distributed over a large population of neurons, which adds another dimension for coding. Nevertheless, our analysis provides important constraints on population coding by characterizing the performance of individual neurons and the diversity of performance levels in a population. A rigorous computational analysis of population coding in field L will require simultaneous recordings from neural populations to characterize interneuronal correlations, because these correlations may significantly impact the performance of the population (Zohary et al., 1994; Romo et al., 2003). Finally, in the present study, individual songs were presented in quiet backgrounds. Another way to extend this study would be to make the task more difficult by adding a masking noise in the background, e.g., a chorus of other birds. Neurometric functions obtained over a range of targets to masker ratios can then be compared with the psychometric functions (Best et al., 2005; Shinn-Cunningham et al., 2006). These experiments should further

challenge the capacity of single cortical neurons in discriminating natural stimuli in more complex environments, e.g., a cocktail party.

References

- Abbott LF, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. *Neural Comput* 11:91–101.
- Bar-Yosef O, Rotman Y, Nelken I (2002) Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. *J Neurosci* 22:8619–8632.
- Berry II MJ, Meister M (1998) Refractoriness and neural precision. *J Neurosci* 18:2200–2211.
- Best V, Ozmeral E, Gallun FJ, Sen K, Shinn-Cunningham BG (2005) Spatial unmasking of birdsong in human listeners: energetic and informational factors. *J Acoust Soc Am* 118:3766–3773.
- Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci* 12:4745–4765.
- Buonomano DV (2000) Decoding temporal information: a model based on short-term synaptic plasticity. *J Neurosci* 20:1129–1141.
- Chacron MJ, Longtin A, Maler L (2001) Negative interspike interval correlations increase the neuronal capacity for encoding time-dependent stimuli. *J Neurosci* 21:5328–5343.
- Cohen MR, Newsome WT (2004) What electrical microstimulation has revealed about the neural basis of cognition. *Curr Opin Neurobiol* 14:169–177.
- Coleman MJ, Mooney R (2004) Synaptic transformations underlying highly selective auditory representations of learned birdsong. *J Neurosci* 24:7251–7265.
- Cynx J (1993) Conspecific song perception in zebra finches (*Taeniopygia guttata*). *J Comp Psychol* 107:395–402.
- David SV, Vinje WE, Gallant JL (2004) Natural stimulus statistics alter the receptive field structure of v1 neurons. *J Neurosci* 24:6991–7006.
- Dayan P, Abbott LF (2001) Theoretical neuroscience: computational and mathematical modeling of neural systems. Cambridge, MA: MIT.
- Doupe AJ, Kuhl PK (1999) Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* 22:567–631.
- Felsen G, Touryan J, Han F, Dan Y (2005) Cortical sensitivity to visual features in natural scenes. *PLoS Biol* 3:e342.
- Furukawa S, Middlebrooks JC (2002) Cortical representation of auditory space: information-bearing features of spike patterns. *J Neurophysiol* 87:1749–1762.
- Grace JA, Amin N, Singh NC, Theunissen FE (2003) Selectivity for conspecific song in the zebra finch auditory forebrain. *J Neurophysiol* 89:472–487.
- Green DM, Swets JA (1966) Signal detection theory and psychophysics. New York: Wiley.
- Gutig R, Sompolinsky H (2006) The tempotron: a neuron that learns spike timing-based decisions. *Nat Neurosci* 9:420–428.
- Hernandez A, Zainos A, Romo R (2000) Neuronal correlates of sensory discrimination in the somatosensory cortex. *Proc Natl Acad Sci USA* 97:6191–6196.
- Hopfield JJ (2004) Encoding for computation: recognizing brief dynamical patterns by exploiting effects of weak rhythms on action-potential timing. *Proc Natl Acad Sci USA* 101:6255–6260.
- Hopfield JJ, Brody CD (2001) What is a moment? Transient synchrony as a collective mechanism for spatiotemporal integration. *Proc Natl Acad Sci USA* 98:1282–1287.
- Lu T, Wang X (2004) Information content of auditory cortical responses to time-varying acoustic stimuli. *J Neurophysiol* 91:301–313.
- Machens CK, Schutze H, Franz A, Kolesnikova O, Stemmler MB, Ronacher B, Herz AV (2003) Single auditory neurons rapidly discriminate conspecific communication signals. *Nat Neurosci* 6:341–342.
- Machens CK, Wehr MS, Zador AM (2004) Linearity of cortical receptive fields measured with natural sounds. *J Neurosci* 24:1089–1100.
- MacLeod K, Backer A, Laurent G (1998) Who reads temporal information contained across synchronized and oscillatory spike trains? *Nature* 395:693–698.
- Major G, Tank D (2004) Persistent neural activity: prevalence and mechanisms. *Curr Opin Neurobiol* 14:675–684.
- Narayan R, Ergun A, Sen K (2005) Delayed inhibition in cortical receptive

- fields and the discrimination of complex stimuli. *J Neurophysiol* 94:2970–2975.
- Narayan R, Grana G, Sen K (2006) Distinct time scales in cortical discrimination of natural sounds in songbirds. *J Neurophysiol* 96:252–258.
- Nelken I, Chechik G, Mrisic-Flogel TD, King AJ, Schnupp JW (2005) Encoding stimulus information by spike numbers and mean response time in primary auditory cortex. *J Comput Neurosci* 19:199–221.
- Osborne LC, Bialek W, Lisberger SG (2004) Time course of information about motion direction in visual area MT of macaque monkeys. *J Neurosci* 24:3210–3222.
- Parker AJ, Newsome WT (1998) Sense and the single neuron: probing the physiology of perception. *Annu Rev Neurosci* 21:227–277.
- Perez-Orive J, Mazor O, Turner GC, Cassenaer S, Wilson RI, Laurent G (2002) Oscillations and sparsening of odor representations in the mushroom body. *Science* 297:359–365.
- Phillips DP, Hall SE, Boehnke SE (2002) Central auditory onset responses, and temporal asymmetries in auditory perception. *Hear Res* 167:192–205.
- Purushothaman G, Bradley DC (2005) Neural population code for fine perceptual decisions in area MT. *Nat Neurosci* 8:99–106.
- Ratnam R, Nelson ME (2000) Nonrenewal statistics of electrosensory afferent spike trains: implications for the detection of weak sensory signals. *J Neurosci* 20:6672–6683.
- Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W (1997) *Spikes: exploring the neural code*. Cambridge, MA: MIT.
- Romo R, Salinas E (2003) Flutter discrimination: neural codes, perception, memory and decision making. *Nat Rev Neurosci* 4:203–218.
- Romo R, Hernandez A, Zainos A, Salinas E (2003) Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron* 38:649–657.
- Schaette R, Gollisch T, Herz AV (2005) Spike-train variability of auditory neurons in vivo: dynamic responses follow predictions from constant stimuli. *J Neurophysiol* 93:3270–3281.
- Schnupp JW, Hall TM, Kokelaar RF, Ahmed B (2006) Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *J Neurosci* 26:4785–4795.
- Schreiber S, Fellous JM, Whitmer D, Tiesinga P, Sejnowski TJ (2003) A new correlation-based measure of spike timing reliability. *Neurocomputing* 52–54:925–931.
- Sen K, Theunissen FE, Doupe AJ (2001) Feature analysis of natural sounds in the songbird auditory forebrain. *J Neurophysiol* 86:1445–1458.
- Shamir M, Sen K, Colburn HS (2007) Temporal coding of time varying stimuli. *Neural Comput*, in press.
- Sharpee TO, Sugihara H, Kurgansky AV, Rebrik SP, Stryker MP, Miller KD (2006) Adaptive filtering enhances information transmission in visual cortex. *Nature* 439:936–942.
- Shinn-Cunningham BG, Best V, Dent ML, Gallun FJ, McClaine EM, Narayan R, Ozmeral E, Sen K (2006) Behavioral and neural identification of bird-song under several masking conditions. *Proceedings of the International Symposium on Hearing*, pp 164–170. Cloppenburg, Germany.
- Sompolinsky H, Yoon H, Kang K, Shamir M (2001) Population coding in neuronal systems with correlated noise. *Phys Rev E Stat Nonlin Soft Matter Phys* 64:051904.
- Strong SP, Koberle R, de Ruyter van Steveninck R, Bialek W (1998) Entropy and information in neural spike trains. *Phys Rev Lett* 80:197–200.
- Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 20:2315–2331.
- van Rossum MC (2001) A novel spike distance. *Neural Comput* 13:751–763.
- Victor JD, Purpura KP (1996) Nature and precision of temporal coding in visual cortex: a metric-space analysis. *J Neurophysiol* 76:1310–1326.
- Victor JD, Purpura KP (1997) Metric-space analysis of spike-trains: theory, algorithms and application. *Network: computation in neural systems* 8:127–164.
- Wehr M, Zador AM (2003) Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* 426:442–446.
- Woolley SM, Fremouw TE, Hsu A, Theunissen FE (2005) Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat Neurosci* 8:1371–1379.
- Wright BD, Sen K, Bialek W, Doupe AJ (2002) Spike timing and the coding of naturalistic sounds in a central auditory area of songbirds. In: *Advances in neural information processing systems*. Cambridge, MA: MIT.
- Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370:140–143.