



# Millisecond-Scale Motor Encoding in a Cortical Vocal Area

Claire Tang<sup>1,2</sup>, Diala Chehayeb<sup>2</sup>, Kyle Srivastava<sup>3</sup>, Ilya Nemenman<sup>2,4</sup>, Samuel J. Sober<sup>2\*</sup>

**1** Neuroscience Graduate Program, University of California, San Francisco, San Francisco, California, United States of America, **2** Department of Biology, Emory University, Atlanta, Georgia, United States of America, **3** Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, Georgia, United States of America, **4** Department of Physics, Emory University, Atlanta, Georgia, United States of America

## Abstract

Studies of motor control have almost universally examined firing rates to investigate how the brain shapes behavior. In principle, however, neurons could encode information through the precise temporal patterning of their spike trains as well as (or instead of) through their firing rates. Although the importance of spike timing has been demonstrated in sensory systems, it is largely unknown whether timing differences in motor areas could affect behavior. We tested the hypothesis that significant information about trial-by-trial variations in behavior is represented by spike timing in the songbird vocal motor system. We found that neurons in motor cortex convey information via spike timing far more often than via spike rate and that the amount of information conveyed at the millisecond timescale greatly exceeds the information available from spike counts. These results demonstrate that information can be represented by spike timing in motor circuits and suggest that timing variations evoke differences in behavior.

**Citation:** Tang C, Chehayeb D, Srivastava K, Nemenman I, Sober SJ (2014) Millisecond-Scale Motor Encoding in a Cortical Vocal Area. *PLoS Biol* 12(12): e1002018. doi:10.1371/journal.pbio.1002018

**Academic Editor:** James Ashe, University of Minnesota, United States of America

**Received:** April 28, 2014; **Accepted:** October 24, 2014; **Published:** December 9, 2014

**Copyright:** © 2014 Tang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are provided in the Supporting Information.

**Funding:** This work was supported by US National Institutes of Health grants R90DA033462 (CT), P30NS069250 (SJS), R01NS084844 (SJS), and R01DC006636 (SJS); National Science Foundation grant IOS-1208126 (IN); and James S. McDonnell Foundation grant 220020321 (IN). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abbreviations:** LMAN, lateral magnocellular nucleus of the anterior nidopallium; NSB, Nemenman-Shafee-Bialek; RA, robust nucleus of the arcopallium.

\* Email: samuel.j.sober@emory.edu

## Introduction

The relationship between patterns of neural activity and the behaviorally relevant parameters they encode is a fundamental problem in neuroscience. Broadly speaking, a neuron might encode information in its spike rate (the total number of action potentials produced per unit time) or in the fine temporal pattern of its spikes. In sensory systems as diverse as vision, audition, somatosensation, and taste, prior work has demonstrated that information about stimuli can be encoded by fine temporal patterns, in some cases where no information can be detected in a rate code [1–11]. This information present in fine temporal patterns might be decoded by downstream areas to produce meaningful differences in perception or behavior.

However, in contrast to the extensive work on temporal coding in sensory systems, the timescale of encoding in forebrain motor networks has not been explored. It is therefore unknown whether the precise temporal coding of sensory feedback could influence spike timing in motor circuits during sensorimotor learning or whether millisecond-scale spike timing differences in motor networks could result in differences in behavior. Although many studies have shown that firing rates can predict variations in motor output [12–14], to our knowledge no studies have examined whether different spiking patterns in cortical neurons evoke different behavioral outputs even if the firing rate remains the same.

The songbird provides an excellent model system for testing the hypothesis that fine temporal patterns in cortical motor systems can encode behavioral output. Song acoustics are modulated on a broad range of time scales, including fast modulations on the order of 10 ms [15,16]. Vocal patterns are organized by premotor neurons in vocal motor cortex (the robust nucleus of the arcopallium [RA]) (Figure 1a), which directly synapse with motor neurons innervating the vocal muscles [14,15,17]. Bursts of action potentials in RA (Figure 1b) are precisely locked in time to production of vocal gestures (“song syllables”), with millisecond-scale precision, suggesting that the timing of bursts is tightly controlled [18]. Similarly, the ensemble activity of populations of RA neurons can be used to estimate the time during song with approximately 10 ms uncertainty [15]. However, although these prior studies demonstrate that the timing of bursts is tightly aligned to the timing of song syllables, it is unknown how variations in the temporal pattern of spikes might encode the trial-by-trial modulations in syllable acoustics known to underlie vocal plasticity [19]. Significantly, biomechanical studies have shown that vocal muscles in birds initiate and complete their force production within a few milliseconds of activation (far faster than in most mammalian skeletal muscles), suggesting that RA’s downstream targets can transduce fine temporal spike patterns into meaningful differences in behavior [20,21]. However, while it is clear that trial-by-trial variation in spike counts within a 40 ms time window can predict variations in the acoustics of individual song syllables

## Author Summary

A central question in neuroscience is how neurons use patterns of electrical events to represent sensory information and control behavior. Neurons might use two different codes to transmit information. First, signals might be conveyed by the total number of electrical events (called “action potentials”) that a neuron produces. Alternately, the timing pattern of action potentials, as distinct from the total number of action potentials produced, might be used to transmit information. Although many studies have shown that timing can convey information about sensory inputs, such as visual scenery or sound waveforms, the role of action potential timing in the control of complex, learned behaviors is largely unknown. Here, by analyzing the pattern of action potentials produced in a songbird’s brain as it precisely controls vocal behavior, we demonstrate that far more information about upcoming behavior is present in spike timing than in the total number of spikes fired. This work suggests that timing can be equally (or more) important in motor systems as in sensory systems.

[14,22], it is unknown whether the precise timing of spikes within bursts might be an even better predictor of vocal motor output than spike counts.

To quantify the temporal scale of encoding in the vocal motor system, we adapted well-established mathematical tools that have previously been applied to measure information transfer in sensory systems. First, we used a spike train distance metric to quantify the differences between pairs of spike trains produced during different renditions of individual song syllables and a classification scheme to quantify whether distance metrics based on rate or timing yielded the best prediction of acoustic output [23,24]. Second, we used model-independent information theoretic methods to compute the mutual information between spike trains and acoustic features of vocal behavior [8,10]. Crucially, both techniques measure information present in the neural activity at different timescales, allowing us to quantify the extent to which spike timing in motor cortex predicts upcoming behavior.

## Results

We collected extracellular recordings from projection neurons in vocal motor area RA in songbirds (Figure 1a). In total, we analyzed 34 single-unit cases and 91 multiunit cases, where a “case” is defined as a neural recording being active prior to the production of a syllable (Figure 1b), as explained in Materials and Methods. The number of trials (syllable iterations) recorded in each case varied from 51 to 1,003 (median 115, mean 192.4). Iterations of each song syllable were divided into groups based on acoustic similarity (“behavioral groups”) (Figure 1c and 1d), and information-theoretic analyses were used to quantify whether spike timing conveys significant information about upcoming motor output, as schematized in Figure 1e.

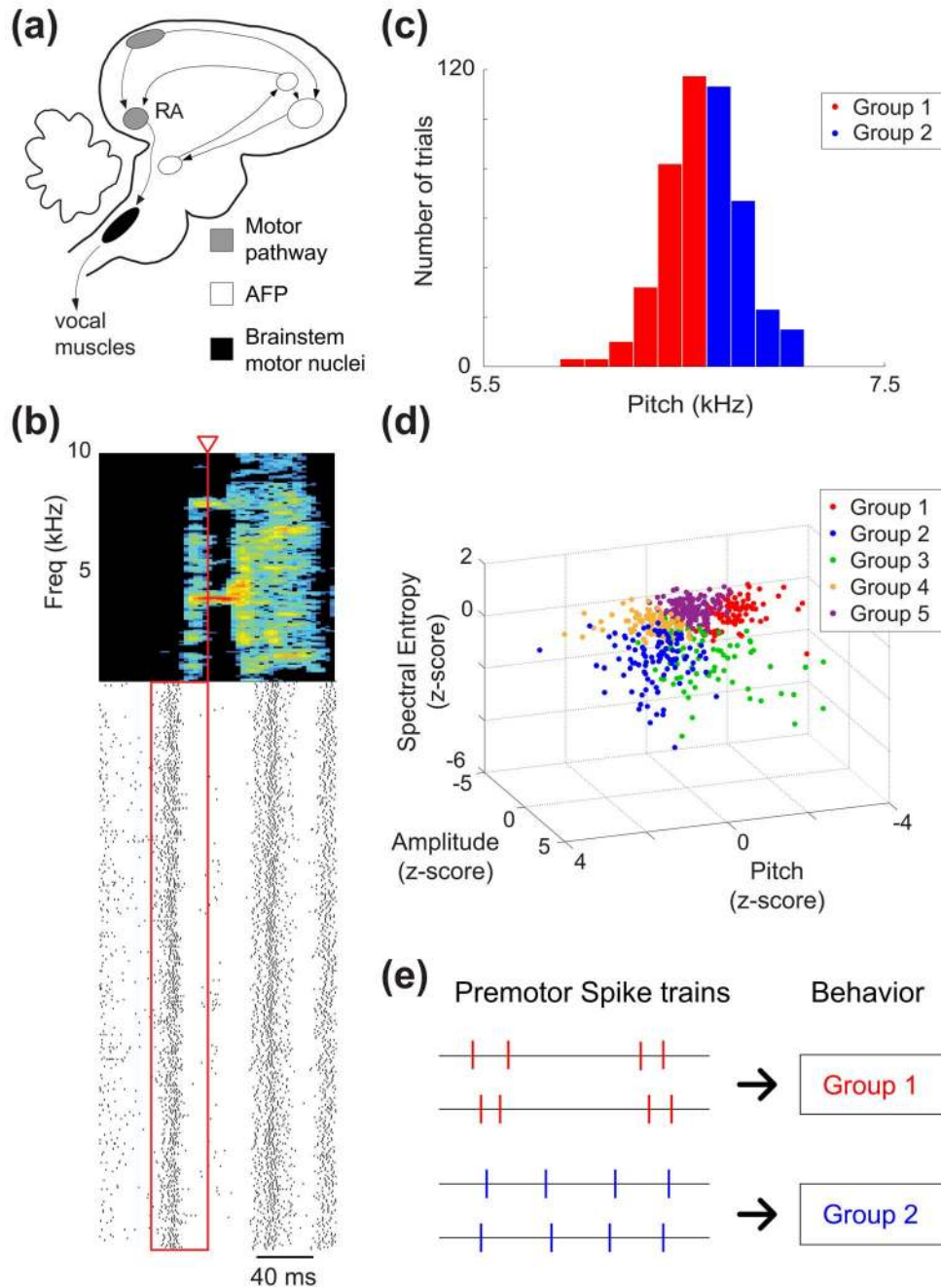
### Metric-Space Analysis

We first used a version of the metric-space analysis established by Victor and Purpura to compare the information conveyed by spike rate and spike timing [24,25]. As described in Materials and Methods, this analysis quantifies how mutual information between neural activity and motor output depends on a cost parameter  $q$ , which quantifies the extent to which spike timing (as opposed to spike number) contributes to the dissimilarity, or “distance,”

between spike trains (Figure 2a). The distance between two spike trains is computed by quantifying the cost of transforming one spike train into the other. Here, parameter  $q$ , measured in  $\text{ms}^{-1}$ , quantifies the relative cost of changing spike timing by 1 ms, as compared to the fixed cost of 1.0 for adding or subtracting a spike. Spike train distances are then used to classify iterations of each song syllable into behavioral groups, and the performance of the classifier  $I(G^P, G)$  is used to quantify the mutual information between neural activity and vocal output. Figure 2b shows a representative “rate case,” where  $q_{\max} = 0$  (that is, information is maximized at  $q = 0$ , where spike train distances are computed based solely on spike counts). As  $q$  increases, the performance of the classifier decreases from its maximal value. This means that the best discrimination between behavioral groups (Figure 1c and 1d) occurs when only spike counts are used in calculating the distances between pairs of spike trains. In contrast, Figure 2c illustrates a “temporal case.” In temporal cases, mutual information between neural activity and vocal motor output reaches its peak when  $q > 0$ . This indicates that there is better discrimination when spike timings are taken into consideration. Note that in the case shown in Figure 2c, the rate code does not provide significant information about behavioral output (empty symbol at  $q = 0$ ).

Across all analyses in cases where information was significant at any value of  $q$ , including cases where  $q_{\max} = 0$ , the median value of  $q_{\max}$  was 0.3, suggesting a high prevalence of temporal cases. Figure 2d shows the prevalence of rate cases and temporal cases in our dataset. As described in Materials and Methods, we assigned the iterations of each song syllable to behavioral groups based either on a single acoustic parameter (e.g., pitch, Figure 1c) or using multidimensional clustering (3D acoustics) (Figure 1d). The different grouping techniques yielded similar results. When syllable acoustics were grouped by clustering in a three-dimensional parameter space (Figure 2d, blue bars), the fraction of temporal cases was significantly greater than the fraction of rate cases (blue asterisk;  $p < 10^{-8}$ , z-test for proportions). Similarly, temporal cases significantly outnumbered rate cases when acoustics were grouped using only a single parameter (pitch, amplitude, or spectral entropy, shown by green, yellow, and red asterisks respectively;  $p < 10^{-8}$ ). Note that in approximately 25% of cases (between 31 and 36 out of 125 cases across the four analyses shown in Figure 2d) these analyses did not yield a significant value of  $I(G^P, G; q)$  for any value of  $q$ ; the fractions in Figure 2d therefore do not sum to unity. In such cases variations in a neuron’s pattern of neural activity during a particular syllable are not predictive of variations in the particular acoustic parameter being analyzed. Furthermore, note that an alternate version of this analysis in which the spike train distance measurement was not normalized by the total number of spikes (see Materials and Methods) yielded nearly identical results, as shown in Figure S1. Additionally, we asked whether the proportions of temporal cases shown in Figure 2d were significantly greater than chance by randomizing the spike times in each trial (Poisson test; Materials and Methods). This analysis revealed a significant proportion of temporal cases when vocal acoustics were measured by multidimensional clustering (3D acoustics,  $p < 0.05$  after Bonferroni correction for multiple comparisons indicated by cross in Figure 2d), but the same measure fell short of significance when the three acoustic parameters were considered individually ( $p = 0.06$ – $0.24$  after Bonferroni correction).

To measure the maximum information available from the metric-space analysis, we computed  $\bar{I}_{\max}$ , the average peak information available across all cases (see Materials and Methods). Across all metric-space analyses  $\bar{I}_{\max}$  was 0.10 bits out of a possible 1.0 bit. As discussed below, this value suggests that additional information might be available in higher-level spike train features



**Figure 1. Neural data and spike train analysis.** (a) The song system consists of two pathways, the direct motor pathway and the anterior forebrain pathway (AFP). Neurons in premotor nucleus RA project to brainstem motor neurons that innervate the vocal muscles. (b) Spike trains recorded from a single RA neuron. Spectrogram of a single song syllable at top shows the acoustic power (color scale) at different frequencies as a function of time. Each tick mark (bottom) represents one spike and each row represents one iteration of the syllable. We analyzed spikes produced in a 40 ms premotor window (red box) prior to the time when acoustic features were measured (red arrowhead). (c) Syllable iterations divided into categories (“behavioral groups”) based on a single acoustic parameter. Here, iterations of a song syllable were divided into two groups ( $N=2$ ; see Materials and Methods) based on fundamental frequency (“pitch”). (d) Syllable iterations divided into  $N=5$  groups by k-means clustering in a three-dimensional acoustic parameter space. (e) We asked whether spike trains could be used to predict differences in behavior. Specifically, our analysis quantifies the extent to which differences in spike timing can discriminate the behavioral group from which the trial was drawn. This is shown in the schematic, in which differences in spike timing contain information about behavioral group even if spike counts (four spikes in this example) are identical across trials.

doi:10.1371/journal.pbio.1002018.g001

that cannot be captured by metric-space analyses. Additionally, since the proportion of rate and temporal cases did not differ significantly when computed from single- or multiunit data ( $p >$

0.07 in all cases; z-tests for proportions), we combined data from both types of recording in this as well as subsequent analyses. The similarity between the single- and multiunit datasets likely results

from multiunit recordings in this paradigm only reflecting the activity of a single or a very small number of neurons, as discussed previously [14]. Finally, the results of the metric-space analysis were not sensitive to the number of behavioral groups used to classify the iterations of each song syllable. Although our primary analysis uses two behavioral groups (Figures 1c and 2), as shown in Table 1 we found a similar prevalence of rate and temporal cases when the trials were divided into three, five (Figure 1d), or eight groups.

Our metric-space analysis therefore indicates that in most RA neurons, taking the fine temporal structure of spike trains into account provides better predictions of trial-by-trial variations in behavior than an analysis of spike rate alone (asterisks, Figure 2d). Furthermore, at least when vocal outputs are grouped in three-dimensional acoustic space, spike timing can predict vocal acoustics significantly more frequently than would be expected from chance (cross, Figure 2d). However, it remains unclear whether spike timing can provide information about single acoustic parameters (beyond the 3-D features). Finally, since not all cases were temporal, it is also unclear how important this information is on average, rather than in special cases. Answering these questions necessitates the direct method of calculating information, as described below.

### Direct Method of Calculating Information

In the metric-space analysis, not all cases were classified as temporal. Further, when behavior was grouped by a single acoustic parameter rather than in multidimensional acoustic space, the number of temporal cases was not significantly larger than by chance (Figure 2d, green, yellow, and red plots). Thus it still remains unclear to what extent spike timing is important to this system overall, rather than in particular instances. Additionally, a drawback of metric-space analyses is that they assume that a particular model (metric) of neural activity is the correct description of neural encoding. As discussed more fully in Materials and Methods, metric-space approaches therefore provide only a lower bound on mutual information [23,25]. Put another way, metric-space analyses assume that the differences between spike trains can be fully represented by a particular set of parameters, which in our case include the temporal separation between nearest-neighbor spike times (Figure 2a). However, if information is contained in higher order aspects of the spike trains that cannot be captured by these parameters (e.g., patterns that extend over multiple spikes), then metric-space analyses can significantly underestimate the true information contained in the neural code. We therefore estimated the amount of information that can be learned about the acoustic group by directly observing the spiking pattern at different temporal resolutions (Figure 3a), without assuming a metric structure, similar to prior approaches in sensory systems [8,10]. We used the Nemenman-Shafee-Bialek (NSB) estimator to quantify the mutual information [26,27]. As described in Materials and Methods, this technique provides minimally biased information estimates, quantifies the uncertainty of the calculated information, and typically requires substantially less data for estimation than many other direct estimation methods [26]. Nevertheless, the NSB technique requires significantly larger datasets than metric-space methods. We therefore directly computed mutual information using the subset (41/125) of cases where the recordings were long enough to gather sufficient data to be analyzed with this method.

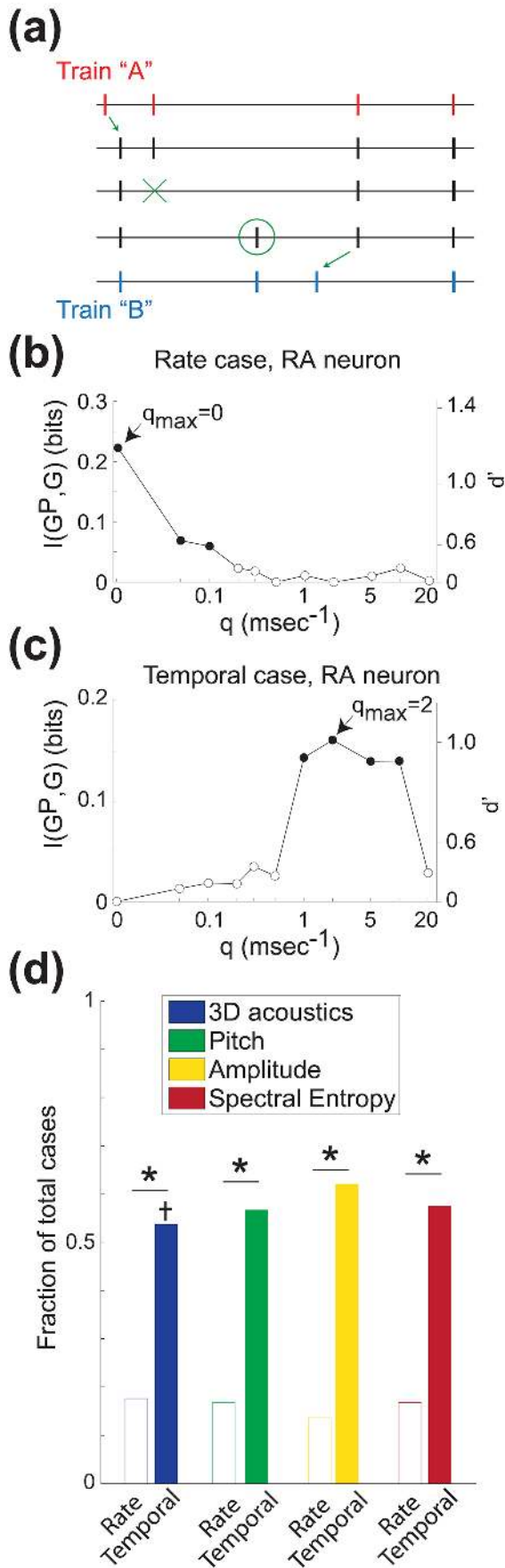
We found that the mutual information between neural activity and vocal behavior rose dramatically as temporal resolution increased. As shown in Figure 3b, when averaged across all 41 cases analyzed using the NSB technique, mutual information was

relatively low when only spike counts were considered (i.e., for  $dt = 40$  ms). Across the four methods of grouping trials based on syllable acoustics, mutual information between spike counts and acoustic output ranged from 0.007 to 0.017 bits (with standard deviations of  $\sim 0.012$ ), which is not significantly different from zero given that all mean information estimates were within  $\sim 1$  SD of zero bits. If information about motor output were represented only in spike counts within the 40 ms premotor window, then mutual information at  $dt < 40$  would be equal to that found at  $dt = 40$  (dashed lines in Figure 3b); note that this is true despite the increase in word length at smaller  $dt$  [8,10]. However, in all analyses mutual information increased as time bin size  $dt$  decreased and reached a maximum value at  $dt = 1$  ms, the smallest bin size (and thus greatest temporal resolution) we could reliably analyze. At 1 ms resolution, mutual information ranged from 0.131 to 0.188 (with standard deviations of  $\sim 0.06$ ) bits across the four analyses performed. These values of mutual information correspond to  $d'$  values near zero at  $dt = 40$  ms and to  $d'$  values between 0.9 and 1.1 at 1-ms resolution (Figure 3b, right-hand axis). Additionally, although Figure 3b shows the results of analyzing data from single-unit and multiunit recordings together, we found very similar results when single- and multiunit data were considered separately (see Figure S2 and Data S1 and S2). These results indicate that far more information about upcoming vocal behavior is available at millisecond timescales and suggest that small differences in spike timing can significantly influence motor output. Therefore, although in some individual cases more information may be available from a rate code in the metric-space analysis (Figure 2d, empty bars), across the population of RA neurons much more information is present in millisecond-scale spike timing. Similarly, note that although in the direct information calculations (Figure 3) mutual information is averaged across all neural recordings, the information at different timescales varied across different neurons (e.g., information at  $dt = 5$  ms in some recordings was greater than information at  $dt = 1$  ms in other recordings). The low mutual information present at  $dt = 40$  ms, for example, therefore reflects the fact that datasets with higher (relative to other datasets) information in the spike count are greatly outnumbered by cases with very low information at  $dt = 40$ .

We performed further analysis to investigate whether the information present at  $dt = 1$  ms reflects differences in burst onset times or differences in the pattern of spikes within bursts (either of which could account for our results). To do so, we performed an alternate analysis (see Materials and Methods, “Inter-Spike Interval Analysis”) in which we calculated the mutual information conveyed by sequences of inter-spike intervals (at a temporal resolution of 1 ms) rather than by absolute spike times. Representing neural activity in this way removes all information about burst onset time and instead quantifies only the mutual information between inter-spike intervals and the acoustic output. In this alternate analysis, information estimates ranged from 0.148 to 0.172 (with standard deviations of  $\sim 0.06$ ) bits across the four different behavioral groupings, and thus were nearly identical to those obtained at  $dt = 1$  ms in our primary analysis. This finding suggests that the information contained in spike timing patterns is carried primarily through the structure of spike timing within bursts rather than by burst onset times.

The results shown in Figure 3 demonstrate that millisecond-scale differences in spike timing can encode differences in behavior. To highlight these timing differences, we examined particular “words” (spike patterns) and considered how different timing patterns could predict vocal acoustics. Figure 4a and 4b each show eight different words from two different single-unit





**Figure 2. Metric-space analysis reveals temporal coding in the vocal motor system.** (a) The distance between example spike trains "A" and "B" is the sum of the fixed costs of adding and subtracting spikes (green circle and X, respectively) and the cost of changing spike timing (green arrows), which is parameterized by the constant  $q$  (see Materials and Methods). (b) Representative rate case. Information  $I(G^P, G)$  about upcoming vocal behavior is maximized when  $q=0$ , indicating a rate code. Filled circles, information significantly greater than zero; empty circles, nonsignificant values. In this case, acoustically similar syllable renditions ("behavioral groups") were grouped by spectral entropy. (c) Representative temporal case. Here information is maximized when  $q>0$ , indicating a temporal code. Note that there is no information in the spike count (unfilled circle at  $q=0$ ). In this example, syllables were grouped by pitch. Right-hand vertical axes in (b) and (c) shows information values converted into  $d'$  units (note nonlinear scale). (d) Prevalence of rate and temporal cases. For each acoustic grouping, the proportion of temporal cases is significantly greater than the proportion of rate cases (asterisks,  $p<10^{-8}$ , z-tests for proportions). Proportions of rate and temporal cases do not differ significantly across the four behavioral groupings. Furthermore, the proportion of temporal cases is significantly greater than that expected by chance for 3-D acoustics (cross,  $p<0.05$ , Poisson test with Bonferroni correction). In all analyses shown, the maximum possible information is 1 bit ( $N=2$  behavioral groups; see Materials and Methods), which corresponds to perfect discrimination between groups. doi:10.1371/journal.pbio.1002018.g002

recordings, color-coded according to the behavioral group in which each word appears most frequently. All words shown in Figure 4 contain the same number of spikes, and thus are identical at the time resolution of  $dt = 40$  ms (Figure 3a). In the example shown in Figure 4a, a distinct set of spike timing patterns predicts the occurrence of low-pitched (group 1) or high-pitched (group 2) syllable renditions. In Figure 4b, behavioral groupings are performed in the three-dimensional acoustic space and similarly show that distinct spike timing patterns can predict vocal acoustics. In some cases, the timing patterns associated with behavioral groups share intuitive features. For example, the words associated with higher pitch in Figure 4a (blue boxes in grid) have shorter inter-spike intervals (but the same total number of spikes) compared with words associated with lower pitch (Figure 4a, red boxes), suggesting that fine-grained interval differences drive pitch variation. However, in other cases (e.g., Figure 4b) no such common features were apparent. Future studies incorporating realistic models of motor neuron and muscle dynamics are therefore required to understand how the precise timing patterns in RA can evoke differences in vocal behavior.

**Comparing Information Estimates across Analyses**

We compared the maximum information available from the metric-space analysis (see Materials and Methods), which is  $I_{max} = 0.10$ , to the information available at the smallest  $dt = 1$  ms in the direct information calculation,  $MI_{NSB} = 0.16$  bits. Reassuringly, the peak information available from the direct method is of the same order of magnitude but somewhat larger than that computed independently in the metric-space analysis. This finding points at consistency between the methods and yet suggests that additional information may be present in higher order spike patterns that cannot be accounted for by a metric-space analysis, namely in temporal arrangements of three or more spikes. (Note, however, that in a small number of cases we were unable to compute mutual information at the finest timescales using the direct method, possibly leading to small biases in our estimates of mutual information at  $dt = 1$  ms; see Materials and Methods). Similarly, a common technique in metric-space analysis is to estimate the "optimal time scale" of encoding as  $1/q_{max}$  (although other authors suggest that such estimates may be highly imprecise

**Table 1.** Effect of dividing trials for each case into a different number of behavioral groups.

<i>N</i>	3-D Acoustics	Pitch	Amplitude	Spectral Entropy
	R, T in %	R, T in %	R, T in %	R, T in %
2	17.6, 53.6 <sup>a</sup>	16.8, 56.0 <sup>a</sup>	13.6, 61.6 <sup>a</sup>	16.8, 56.8 <sup>a</sup>
3	26.4, 53.6 <sup>a</sup>	35.2, 45.6	20.8, 48.0 <sup>a</sup>	20.0, 57.6 <sup>a</sup>
5	20.0, 54.4 <sup>a</sup>	26.4, 53.6 <sup>a</sup>	24.0, 50.4 <sup>a</sup>	31.2, 59.2 <sup>a</sup>
8	23.2, 54.4 <sup>a</sup>	24.8, 49.6 <sup>a</sup>	18.4, 61.6 <sup>a</sup>	20.0, 59.2 <sup>a</sup>

Numbers in the table are the percentages of total cases (R) that are rate cases (T) and temporal cases, respectively.

<sup>a</sup>Instances where the proportion of temporal cases is significantly greater than the proportion of rate cases.

doi:10.1371/journal.pbio.1002018.t001

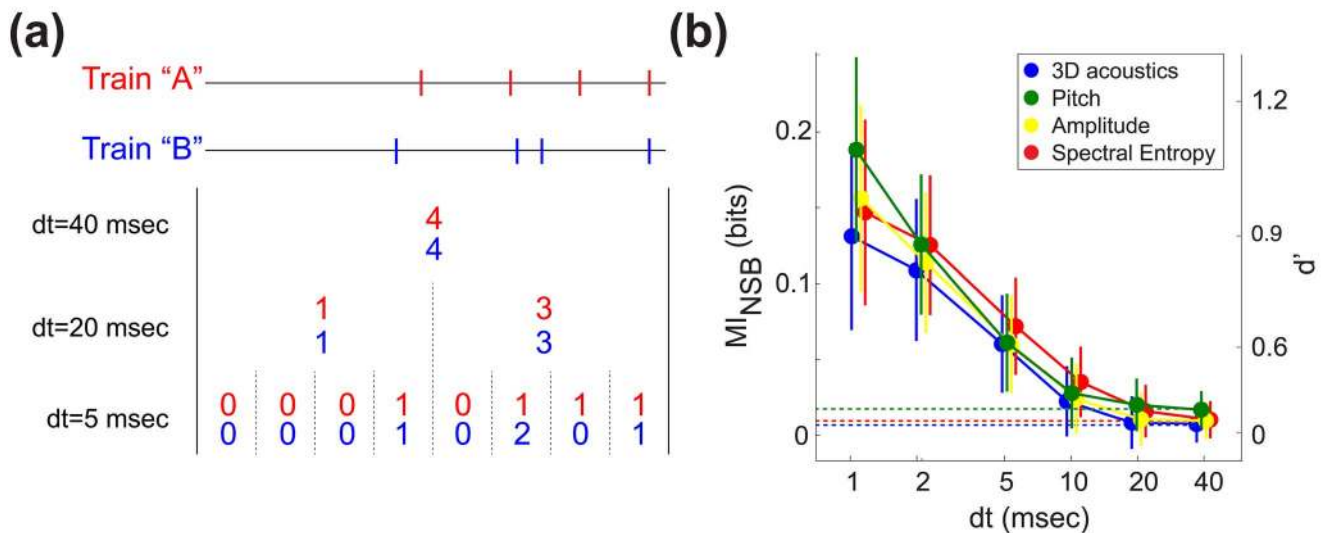
[25]). In our dataset, the median value of  $q_{max}$  was  $0.3 \text{ ms}^{-1}$ , suggesting that spike timing precision is important down to  $1/q_{max} \sim 1 \text{ ms}$ , which is again in agreement with the direct estimation technique.

### Discussion

We computed the mutual information between premotor neural activity and vocal behavior using two well-established computational techniques. A metric-space analysis demonstrated that spike timing provides a better prediction of vocal output than spike rate in a significant majority of cases (Figure 2). A direct computation of mutual information, which was only possible in the subset of recordings that yielded relatively large datasets, revealed that the amount of information encoded by neural activity was maximal at a 1 ms timescale, while the average information available from a rate code was insignificant (Figure 3). It also suggested that

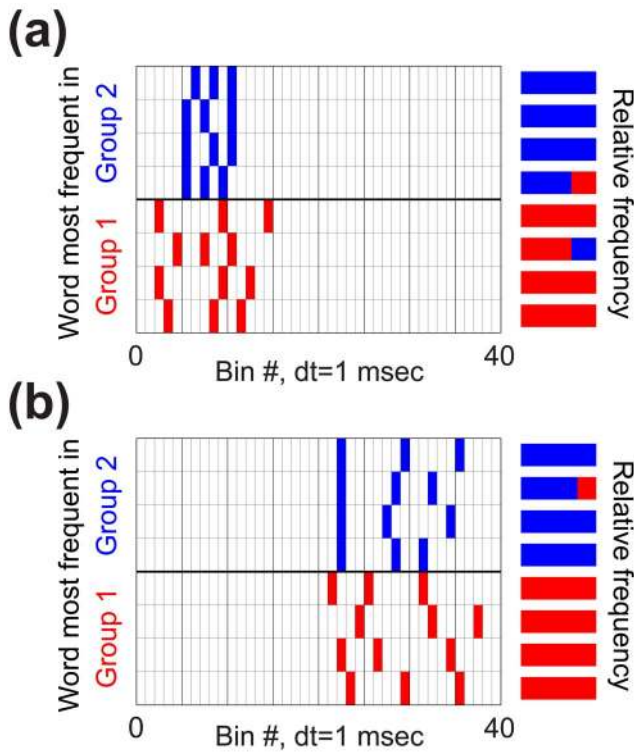
information in the spike trains may be encoded in higher order spike patterns.

Although previous studies have shown that bursts in RA projection neurons are aligned in time to the occurrence of particular song syllables [15,18], ours is the first demonstration to our knowledge that variations in spike timing within these bursts can predict trial-by-trial variations in vocal acoustics. These acoustic variations are thought to underlie vocal learning ability in songbirds. A number of studies have demonstrated that nucleus LMAN (the lateral magnocellular nucleus of the anterior nidopallium), the output nucleus of the anterior forebrain pathway (AFP) and an input to RA (Figure 1a), both generates a significant fraction of vocal variability and is required for adaptive vocal plasticity in adult birds [28–30]. A significant question raised by our results therefore concerns the extent to which LMAN inputs can alter the timing of spikes in RA. Recent work has shown that spike timing patterns in LMAN neurons encode the time during song [31]. Future studies



**Figure 3. Direct calculation of information reveals more information at finer temporal resolution.** (a) The 40 ms-long spike train prior to each song syllable was converted into “words” with different time resolutions ( $dt$ ), where the symbols within each word represent the number of spikes per bin. At  $dt = 40 \text{ ms}$ , two spike trains (“A” and “B”) from our dataset are both represented by the same word (4). However, when  $dt$  decreases to 5 ms, the trains A and B were represented by different words ([0 0 0 1 0 1 1 1] and [0 0 0 1 0 2 0 1], respectively). We used the NSB entropy estimation technique to directly compute the mutual information between the distribution of words and vocal acoustics at different temporal resolutions (see Materials and Methods). (b) Mutual information ( $MI_{NSB}$ ) increases as  $dt$  decreases. There is close to no information in the spike count,  $dt = 40$ . Right-hand vertical axis shows information values converted into  $d'$  units. Error bars represent 1 standard deviation (SD) of the information estimate. Here, the number of acoustic groups is 2 and the maximum possible information is therefore 1 bit. Dashed lines indicate mutual information at the 40 ms timescale and illustrate the mutual information expected at  $dt < 40$  if no information were present at faster timescales (i.e., from a rate code; see text).

doi:10.1371/journal.pbio.1002018.g003



**Figure 4. Spike patterns within bursts predict vocal acoustics.** Each grid shows eight “words” at time resolution  $dt=1$  ms (see Figure 3a). Here we consider words with equal numbers of spikes (three). Rows represent different words, columns represent characters within a word, and boxes are filled when a spike is present. Words are color-coded according to which behavioral group they appear in most frequently, with words appearing more often in groups 1 and 2 shown in red and blue, respectively. Colored bars at right show the relative frequency with which each word appears in group 1 or 2, for example a solid red bar indicates a word that only occurs in behavioral group 1. Data in (a) are from the same single-unit recording shown in Figure 1b, with behavioral groups determined by pitch (Figure 1c). Data in (b) are from a different single-unit recording with behavioral grouping in 3-D acoustic space. Note that although this figure illustrates subsets of observed words, mutual information is always computed over the full distribution of all words.  
doi:10.1371/journal.pbio.1002018.g004

might address whether the observed patterns in LMAN spiking can also predict acoustic variations, and lesion or inactivation experiments could quantify changes in the distribution of firing patterns in RA after the removal of LMAN inputs [32].

Our results indicate that spike timing in cortical motor networks can carry significantly more information than spike rates. Equivalently, these findings suggest that limiting the analysis of motor activity to spike counts can lead to drastic underestimates of information. This contrast is illustrated by a comparison of the present analysis and our prior study examining correlations between premotor spike counts and the acoustics of song syllables [14]. In that earlier study, we found that spike rate predicted vocal output in  $\sim 24\%$  of cases, a prevalence similar to the proportion of rate cases observed in the metric-space analysis and far smaller than the prevalence of temporal cases (Figure 2). Similarly, direct computations of mutual information (Figure 3) show that a purely rate-based analysis would detect only a small fraction of the information present in millisecond-scale timing. Therefore our central finding—that taking spike timing into account greatly increases the mutual information between neural activity and

behavior—suggests that correlation and other rate-based approaches to motor encoding might in some cases fail to detect the influence of neural activity on behavior.

As shown in Figure 3, we found that spike timing at the 1 ms timescale provides an average of  $\sim 0.16$  bits out of a possible 1.0 bit of information when discriminating between two behavioral groups. While this value is of course less than the maximum possible information, it is important to note that this quantity represents the average information available from a single neuron. A number of studies in sensory systems have demonstrated that ensembles of neurons can convey greater information than can be obtained from single neurons [33]. While our dataset did not include sufficient numbers of simultaneous recordings to address this issue, future analyses of ensemble recordings could test the limits of precise temporal encoding in the motor system.

Temporal encoding in the motor system could also provide a link between sensory processing and motor output. Prior studies have shown that different auditory stimuli can be discriminated on the basis of spike timing in auditory responses [11,34,35], including those in area HVC, one of RA’s upstream inputs [36]. Our results demonstrate that in songbirds, temporally precise encoding is present at the motor end of the sensorimotor loop. Propagating sensory-dependent changes in spike timing into motor circuits during behavior might therefore underlie online changes in motor output in response to sensory feedback [37,38] or serve as a substrate for long-term changes in motor output resulting from spike timing-dependent changes in synaptic strength [19,39–41].

While the existence of precise spike timing is strongly supported for a variety of sensory systems, a lingering question is how downstream neural networks could use the information that is present at such short timescales, and hence whether the animal’s behavior could be affected by the details of spike timing. Although theoretical studies have suggested how downstream neural circuits could decode timing-based spike patterns in sensory systems [42], the general question of whether the high spiking precision in sensing, if present, is an artifact of neuronal biophysics or a deliberate adaptation remains unsettled [43].

In motor systems, in contrast, spike timing differences could be “decoded” via the biomechanics of the motor plant, thereby transforming differences in spike timing into measurable differences in behavior. In a wide range of species [44–47], the amplitude of muscle contraction can be strongly modulated by spike timing differences in motor neurons (i.e., neurons that directly innervate the muscles) owing to strong nonlinearities in the transformation between spiking input and force production in muscle fibers. Furthermore, biomechanical studies have shown that vocal muscles in birds have extraordinarily fast twitch kinetics and can reach peak force production in less than 4 ms after activation [20,21], suggesting that the motor effectors can transduce millisecond-scale differences in spike arrival into significant differences in acoustic output. Finally, *in vitro* and modeling studies have quantified the nonlinear properties in the songbird vocal organ, demonstrating that small differences in control parameters can evoke dramatic and rapid transitions between oscillatory states, suggesting again that small differences in the timing of motor unit activation could dramatically affect the acoustics of the song [48,49]. Future studies that model the dynamics of brainstem networks downstream of RA as well as the mechanics of the vocal organ could address how particular spiking patterns in RA (such as those shown in Figure 4) might drive variations in acoustic output.

Our results demonstrate that the temporal details of spike timing, down to 1 ms resolution, carry about ten times as much

information about upcoming motor output compared to what is available from a rate code. This is in marked contrast to sensory coding [8,10], where the information from spike patterns at millisecond resolution is often about double that available from the rate alone. For this reason, the most striking result of our analysis might be that precise spike timing in at least some motor control systems appears to be even more important than in sensory systems. In summary, although future work in both sensory and motor dynamics is needed to fully explicate how differences in spike timing are mapped to behavioral changes, our findings, in combination with previous results from sensory systems, represent the first evidence, to our knowledge, for the importance of millisecond-level spiking precision in shaping behavior throughout the sensorimotor loop.

## Materials and Methods

### Ethics

All procedures were approved by the Emory University Institutional Animal Care and Use Committee.

To measure the information about vocal output conveyed by motor cortical activity at different timescales, we recorded the songs of Bengalese finches while simultaneously collecting physiological data from neurons in RA. We then quantified the acoustics of individual song syllables and divided the iterations of each syllable into “behavioral groups” on the basis of acoustic features such as pitch, amplitude, and spectral entropy. Mutual information was then computed using two complementary techniques. First, we used a metric-space analysis [23] to quantify how well the distance between pairs of spike trains can be used to classify syllable iterations into behavioral groups. Second, we used a direct calculation of mutual information [10,26,27,50] to produce a minimally biased estimate of the information available at different timescales.

### Neural Recordings

Single-unit and multiunit recordings of RA neurons were collected from four adult (>140 days old) male Bengalese finches using techniques described previously [14]. All procedures were approved by the Emory University Institutional Animal Care and Use Committee. Briefly, an array of four or five high-impedance microelectrodes was implanted above RA nucleus. We advanced the electrodes through RA using a miniaturized microdrive to record extracellular voltage traces as birds produced undirected song (i.e., no female bird was present). We used a previously-described spike sorting algorithm [14] to classify individual recordings as single-unit or multiunit. In total, we collected 53 RA recordings (19 single-unit, 34 multiunit), which yielded 34 single-unit and 91 multiunit “cases,” as defined below. Based on the spike waveforms and response properties of the recordings, all RA recordings were classified as putative projection neurons that send their axons to motor nuclei in the brainstem [14,15,51]. A subset of these recordings has been presented previously as part of a separate analysis [14].

### Acoustic Analysis and Premotor Window

We quantified the acoustics of each song syllable as described in detail previously [14]. Briefly, for each syllable we measured vocal acoustics at a particular time (relative to syllable onset) when spectral features were well defined (Figure 1b, red line). Syllable onsets were defined on the basis of amplitude threshold crossings after smoothing the acoustic waveform with a square filter of width

2 ms; therefore the measurement error of our quantification of syllable onset time is on the order of milliseconds. Note that this uncertainty cannot account for our results, since millisecond-scale jitter in syllable onset (and thus burst timing) will decrease, rather than increase, the amount of information present at fine timescales. We quantified the fundamental frequency (which we refer to here as “pitch”), amplitude, and spectral entropy by analyzing the acoustic power spectrum at the specified measurement time during each iteration of a song syllable. We selected these three acoustic features because they capture a large percentage of the acoustic variation in Bengalese finch song [14]. In the example syllable illustrated in Figure 1b (top), the band of power at ~4 kHz is the fundamental frequency. Furthermore, each sound recording was inspected for acoustic artifacts unrelated to vocal production and these trials, which constituted less than 1% of the total data, were discarded to minimize potential measurement error. For each iteration of each syllable, we analyzed spikes within a temporal window prior to the time at which acoustic features were measured. The width of this window was selected to reflect the latency with which RA activity controls vocal acoustics. Although studies employing electrical stimulation have produced varying estimates of this latency [52,53], a single stimulation pulse within RA modulates vocal acoustics with a delay of 15–20 ms [54]. We therefore set the premotor window to begin 40 ms prior to the time when acoustic features were measured and to extend until the measurement time (Figure 1b, red box). This window therefore includes RA’s premotor latency [14,22] and allows for the possibility that different vocal parameters have different latencies.

### Determining Behavioral Groups

While grouping spike trains is straightforward in many sensory studies, where different stimuli are considered distinct groups, we face the problem of continuous behavioral output in motor systems. We took two approaches to binning continuous motor output into discrete classes. First, we considered only a single acoustic parameter and divided the trials into equally sized groups using all of the data. For example, Figure 1c shows trials divided into two behavioral groups based on one parameter (pitch). In addition to pitch, separate analyses also used sound amplitude or spectral entropy to divide trials into groups. In the second approach (which we term “3D acoustics”) (Figure 1d), we used k-means clustering to divide trials into groups. Clustering was performed in the three-dimensional space defined by pitch, amplitude, and entropy, with raw values transformed into z-scores prior to clustering. Note that both approaches allow us to divide the dataset into an arbitrary number of groups (parameter  $N$ , see “Discrimination analysis” below). Our primary analysis divided trials into  $N = 2$  groups since a smaller  $N$  increases statistical power by increasing the number of data points in each group. However, alternate analyses using greater  $N$  yielded similar conclusions (see Results).

### Information Calculation I: Metric-Space Analysis

In previous studies, metric-space analysis has been used to probe how neurons encode sensory stimuli (for a review, see [55]). The fundamental idea underlying this approach is that spike trains from different groups (e.g., spikes evoked by different sensory stimuli) should be less similar to each other than spike trains from the same group (spikes evoked by the same sensory stimulus). In the present study, we adapt this technique for use in the vocal motor system to ask how neurons encode trial-by-trial variations in the acoustic structure of individual song syllables. To do so, we divide the iterations of a song syllable into “behavioral groups”



based on variations in acoustic structure (Figure 1c). We then construct a “classifier” to ask how accurately each spike train can be assigned to the correct behavioral group using a distance metric that quantifies the dissimilarity between pairs of spike trains [24]. As described in detail below, the classifier attempts to assign each trial to the correct behavioral group on the basis of the distances between that trial’s spike train and the spike trains drawn from each behavioral group. Crucially, the distance metric is parameterized by  $q$ , which reflects the importance of spike timing to the distance between two spike trains. This method therefore allows us to evaluate the contribution of spike timing to the performance of the classifier, and thus to the information contained in the spike train about the behavioral group.

**Calculating distances.** The distance metric used in this study,  $D[q]$ , is a normalized version of the distance metric  $D_{spike}[q]$  originally introduced by Victor and Purpura [24,25]. The original metric is defined as the minimal cost of transforming one spike train into a second. There are three elementary steps, insertion (Figure 2a, green circle) and deletion (Figure 2a, green X) of a spike, which have a cost of 1, and shifting a spike (Figure 2a, green arrows), which has a cost that is directly proportional to the amount of time the spike is moved. The proportionality constant,  $q$ , can take on values from 0 to infinity. When  $q=0$ , there is no cost for shifting spikes, and the distance between two spike trains is simply the absolute value of the difference between the number of spikes in each. For  $q>0$ , spike timings matter, and distances are smaller when spike times are similar between the two spike trains. The distance is normalized by dividing by the total number of spikes from both spike trains. The normalized version of the Victor and Purpura distance is more consistent with the assumption that spike trains with the same underlying rate should have smaller distances than spike trains with different rates [25]. However, as described in Results and shown in Figure S1, we obtained nearly identical results in an alternate analysis that used the unnormalized version of this distance metric. Importantly, the time-scale parametric of  $D[q]$  allows us to evaluate the contribution of spike timing to the amount of information transmitted by the neuron about the behavior.

**Classifier-based measurement of mutual information.** To determine the amount of systematic, group-dependent clustering, a decoding algorithm (“classifier”) is used to classify the spike trains into predicted groups based on  $D[q]$ . The performance of the classifier in discriminating between behavioral groups is measured by calculating the mutual information between the actual group and predicted group.

The classifier assigns trials to a predicted group by minimizing the average distance to the group. Given a spike train  $s$ , we calculate the average distance from  $s$  to the spike trains pertaining to a certain group  $G_i$  by:

$$d(s, G_i) = \left[ \langle D[q](s, s')^z \rangle_{s' \text{ from group } G_i} \right]^{1/z} \quad (1)$$

If  $s$  belongs to group  $G_i$ , we exclude the term  $D[q](s, s)$  from the above equation. The trial is classified into the group  $G_i$  that minimizes this average distance, and the resulting information is summarized into a confusion matrix  $C(G_k^p, G_l)$ , which indicates the number of times that trial from group  $G_l$  is assigned to group  $G_k^p$ . The parameter  $z$  determines the geometry of the average, biasing the average to the shortest distances for negative values are emphasizing reducing the distance to outliers for positive values.

This procedure is performed for a range of  $q$  values (0, 0.05, 0.1, 0.2, 0.3, 0.5, 1, 2, 5, 10, and 20  $\text{ms}^{-1}$ ) to produce a set of

confusion matrices, which are normalized into probability matrices  $P(G_k^p, G_l)$  by dividing by the total number of spike trains. Then the performance of the classification can be measured by computing the mutual information,  $I$ , between the actual group and predicted group.

$$I(G^p, G; q) = \sum_{k=1}^N \sum_{l=1}^N p(G_k^p, G_l; q) \log_2 \frac{p(G_k^p, G_l; q)}{p(G_k^p; q)p(G_l; q)} \quad (2)$$

The variable  $N$  in Equation 2 refers to the number of groups each dataset’s trials were divided into. Except where otherwise indicated, we used  $N=2$ . To optimize the performance of the classifier, we maximized mutual information across different values of  $z$  in the range  $-8$  to  $8$  for each value of  $q$ , as described previously [25].

$I_{count}(G^p, G)$  is the information when only spike counts are considered, that is when  $q=0$ , or  $I(G^p, G; 0)$ .  $I_{max}(G^p, G)$  is the maximum value of  $I(G^p, G; q)$ , and the value of  $q$  associated with  $I_{max}(G^p, G)$  is  $q_{max}$ . If  $I(G^p, G; q)$  plateaus, obtaining  $I_{max}(G^p, G)$  at more than one value of  $q$ ,  $q_{max}$  is defined as the smallest of those values.

**Bias correction, “classifier.”** Because there is a component of the classification that is correct by chance, the estimate from Equation 2 can overestimate the true information. This bias can be computationally approximated and subtracted from the original estimate [24]. Concretely, we shuffle the spike trains across groups and then perform the analysis 1,000 times and calculate the average information across these random reassignments. This value is an estimate of the bias and is subtracted from the original estimate. After subtraction, only values above the 95th percentile of the null distribution of  $I$  values are considered significant and negative values are set equal to zero.

**Rate cases versus temporal cases.** We define a “case” as one neural recording (single- or multiunit) that meets an average firing threshold of 1 spike in the 40 ms premotor window before one syllable. We limited our analysis to cases for which at least 50 trials were available. After performing the above analyses on each case, we categorized the cases into “rate cases” and “temporal cases.” Rate cases are when the maximum amount of information occurs for  $q=0$ . For rate cases,  $I_{count}=I_{max}$ , indicating that the best discrimination occurs when only spike counts are considered. For cases where  $q_{max}>0$ , the fine temporal structure of the spike train also contributes to discrimination, which we define as a temporal case. Note that this definition of “rate” coding is dependent on our choice of the duration (40 ms) of the premotor window used in our analysis. Therefore, although our classification of each dataset as either a rate or a temporal case provides a meaningful distinction between information being maximized in the spike count within the 40 ms window versus at a finer timescale, our analysis is not informative about rate coding at longer timescales. Although it is possible that activity at premotor latencies greater than 40 ms can carry more information about behavioral output, this is unlikely given that electrical stimulation of RA evokes changes in syllable acoustics at a latency of  $\sim 15$  ms [54].

To determine whether the proportion of temporal cases,  $\hat{p}_T$ , is significantly greater than chance, we constructed synthetic datasets in which we randomized spike times for each trial in each case (Poisson test). These randomized spike trains had the same number of spikes as our original data. We then performed metric-space analysis in the same manner as before and calculated the proportion of temporal cases across all cases. After generating 1,000 of these synthetic datasets, we found the distribution of  $\hat{p}_T$

under the null hypothesis that spike timings do not encode motor output and asked whether our observed  $\hat{p}_T$  was greater than the 95th percentile of this distribution. Additionally, we performed one-sided z-tests for proportions to ask whether the proportion of temporal cases exceeded the proportion of rate cases.

### Information Calculation II: Direct Method

In addition to the metric-space analysis described above, we also directly calculated the mutual information between song acoustics and neural activity [10]. Whereas metric-space analysis makes strong assumptions about the structure of the neural code, the direct approach is model-independent [10,56]. Specifically, spike train distance metrics assume that spike trains that have spike timings closer to each other are linearly more similar than spike trains whose timings are more different. As with all assumptions, the methods gain extra statistical power if they are satisfied, but they may fail if the assumptions do not hold. The direct method simply considers distinct patterns of spikes at each timescale (which can vary in total spike number, the pattern of inter-spike intervals, burst onset time, etc.), without assigning importance to specific differences. Crucially, direct methods allow us to estimate the true mutual information, whereas the mutual information computed from a metric-space analysis represents only a lower bound on this quantity [3]. However, because the direct method is a model-independent approach that does not make strong assumptions about the neural code, it requires larger datasets to achieve statistical power.

To determine whether there is information about acoustics in the precise timing of spikes, we compared the information between neural activity and behavioral group following discretization of the spike trains at different time resolutions. For a time bin of size  $dt$ , each  $T=40$  ms-long spike train was transformed into a “word” with  $40/dt$  symbols, where different symbols represent the number of spikes per bin. The mutual information is simply the difference between the entropy of the total distribution of words  $H_{T,dt}[\mathbf{R}]$  and the average entropy of the words given the behavioral group  $H_{T,dt}[\mathbf{R}|\mathbf{G}]$ :

$$I_{direct\ T,dt}[\mathbf{R}; \mathbf{G}] = H_{T,dt}[\mathbf{R}] - \langle H_{T,dt}[\mathbf{R}|\mathbf{G}] \rangle_G \quad (3)$$

$I_{direct}$  could be quantified exactly if the true probability distributions  $p(\mathbf{R})$ ,  $p(\mathbf{R}, \mathbf{G})$ , and  $p(\mathbf{R}|\mathbf{G})$  were known:

$$I_{direct} = - \sum_{\mathbf{R}} p(\mathbf{R}) \log_2 p(\mathbf{R}) - \left( - \sum_{\mathbf{R}} \sum_{\mathbf{G}} p(\mathbf{R}, \mathbf{G}) \log_2 p(\mathbf{R}|\mathbf{G}) \right) \quad (4)$$

However, estimating these distributions from finite datasets introduces a systematic error (“limited sample bias” or “under-sampling bias”) that must be corrected [57]. There are several methods to correct for this bias, but most assume that there is enough data to be in the asymptotic sampling regime, where each typical response has been sampled multiple times. As we increase the time resolution of the binning of the spike train, the number of possible neural responses increases exponentially, and we quickly enter the severely undersampled regime where not every “word” is seen many times, and, in fact, only a few words happen more than once (which we term a “coincidence” in the data). We therefore employed the NSB entropy estimation technique [26,27], which can produce unbiased estimates of the entropies in Equation 3 even for very undersampled datasets.

The NSB technique uses a Bayesian approach to estimate entropy. However, instead of using a classical prior, for which all values of the probability of spiking are equally likely, NSB starts with the a priori hypothesis that all values of the entropy are equally likely. This approach has been shown to reliably estimate entropy in the severely undersampled regime (where the number of trials per group is much less than the cardinality of the response distribution) provided that the number of coincidences in that data is significantly greater than one. This typically happens when the number of samples is only about a square root of what would be required to be in the well-sampled regime [26,50].

This method often results in unbiased estimates of the entropy, along with the posterior standard deviation of the estimate, which can be used as an error bar on the estimate [50]. On the other hand, we know that no method can be universally unbiased for every underlying probability distribution in the severely undersampled, square-root, regime [58]. Thus there are many underlying distributions of spike trains for which NSB would be biased. Correspondingly, the absence of bias cannot be assumed and must instead be verified for every estimate, which we do as described below.

*A priori*, we restricted our analysis to cases in which the number of trials was large enough ( $>200$ ) so that the number of coincidences would likely be significantly greater than 1. Of our 125 datasets, 41 passed this size criterion. We emphasize that no additional selection beyond the length of recording was done. Since recording length is unrelated to the neural dynamics, we expect that this selection did not bias our estimates in any way. The NSB analysis was performed using  $N=2$  behavioral groups, since increasing the number of groups greatly decreased the number of coincidences and increased the uncertainty of the entropy estimates (not shown). Additionally, because NSB entropy estimation assumes that the words are independent samples, we verified that temporal correlations in the data are low. To do this, we used NSB to calculate the entropy of four different halves of each dataset: the first half of all trials, the second half, and the two sets of every other trial, where the second set is offset from the first set by one trial. We found that the difference in mean entropy between the first half and second half data was very similar to the difference between the two latter sets. Any temporal correlations in our neural data are therefore very low and thus unlikely to affect entropy estimation, and the information at high spiking precision that we observe cannot just be attributed to modulation occurring on a longer time scale.

To make sure that the NSB estimator is unbiased for our data, we estimated each conditional and unconditional entropy from all available  $N$  samples, and then from  $\alpha N$ ,  $\alpha < 1$ , samples. Twenty-five random subsamples of size  $\alpha N$  were taken and then averaged to produce  $\langle S(\alpha) \rangle$ . We plotted  $\langle S(\alpha) \rangle$  versus  $1/\alpha$  and checked whether all estimates for  $1/\alpha \rightarrow 1$  agreed among themselves within error bars, indicating no empirical sample-size dependent bias [8,10]. At temporal resolutions in the tens of milliseconds, virtually all cases showed no sample size-dependent drift in the entropy estimates, and hence the estimates from full data were treated as unbiased. As the temporal resolution increased to  $dt = 1$ , bias was visible in some cases. However, the bias could often be traced to the rank-ordered distribution of words not matching the expectations of the NSB algorithm. Specifically, some of the most common words occurred much more often than expected from the statistics of the rest of the words. Since NSB uses frequencies of common, well-sampled words to extrapolate to undersampled words, such uncommonly frequent outliers can bias entropy estimation [27]. To alleviate the problem, we followed [8] and partitioned the response

distribution in a way such that the most common word was separated from the rest when it was too frequent (with “too frequent” defined as  $>2\%$  of all words). We use  $p_1$  to denote the frequency of the most common word and  $p_2$  to denote the frequency of all other words. We then used the additivity of entropy,

$$S = p_1 S_1 + p_2 S_2 + S(p_1, p_2) \quad (5)$$

to compute the total entropy by first estimating the entropy of the choice between the most common word and all others,  $S(p_1, p_2)$ , and the entropy of all of the data excluding the most common word,  $S_2$ , independently using the NSB method (the entropy of the single most common word,  $S_1$ , is zero). The error bars were computed by summing the individual error bars in Equation 5 in quadratures. Once the most common word was isolated in the cases in which it was “too frequent,” the resulting entropies were checked again for bias using the subsampling procedure explained above. In the majority of cases in which sample-size dependent bias was detected, this bias was removed through the partition-based entropy correction and the resulting entropy estimates were therefore empirically verified as being unbiased. In other biased cases, there were no uncharacteristically common words, and we could not perform the partition-based entropy correction. Furthermore, at the highest temporal resolutions, there were a few cases that did not have enough coincidences for the NSB algorithm to produce an estimate of entropy at all. These cases were excluded from the following steps. The total fraction of entropy estimates found to be unbiased (i.e., was either unbiased to begin with or bias was successfully corrected using partitioning) was 100% at  $dt = 40$  ms and decreased monotonically to 72% at  $dt = 2$  ms and 58% at  $dt = 1$  ms. Thus, we were unable to correct entropy biases in a minority of entropy measurements at fine timescales. However, as described below (and illustrated in Figure S2), our results were nearly identical when we removed all cases with biased entropy from our dataset, demonstrating that any residual biases cannot account for our results.

We then averaged the mutual information between the spike train and the acoustic group over all cases, weighing contribution of each case by the inverse of its respective posterior variance. The variance of the mean was similarly estimated. As an example, the ten biased cases at  $dt = 1$  for syllables grouped by pitch that could not be corrected had few coincidences and hence large error bars, so that the average value of the inverse variance for these ten cases was 14% of the average inverse variance for the 20 cases that were originally unbiased, or corrected to unbiased. Therefore, since 14% of ten cases is approximately one case, these ten biased cases together contributed about as much as one of the 20 unbiased or bias-corrected cases in the final calculation of the average mutual information. These biased cases thus cannot contribute significantly to bias in the average mutual information. This is true for the other behavioral groupings we performed as well. To confirm this, we further performed an alternate analysis (Figure S2d) in which we excluded the biased cases that could not be corrected from the final calculation of mutual information. This alternate analysis yielded very similar results as the original analysis, with a large increase in mutual information at finer temporal resolution. Our results therefore cannot be ascribed to the inclusion of some cases in which entropy biases could not be corrected. Finally, note that as described above we of necessity excluded from our analysis cases in which no coincidences occurred (since the NSB algorithm estimates entropy by counting the number of coincidences). This occurred in a small fraction of cases (13% of entropy estimates at

$dt = 1$  ms, 5% of estimates at  $dt = 2$  ms,  $<1\%$  of estimates at  $dt = 5$  ms, and 0% at  $dt = 10$  ms or greater). Since cases in which no coincidences occur will likely be those with low mutual information, excluding these cases may have introduced a slight upward bias to our information estimates at fine timescales. However, information in such cases would be likely to have large uncertainties and thus would make insignificant contributions to the weighted average, and furthermore the number of excluded cases at timescales greater than 1 ms (crucially, including  $dt = 2$  ms) is negligible. In any event, these excluded cases cannot account for either the non-zero information at fine timescales or for the dramatic increase in information observed as temporal resolution increases.

**Inter-spike interval analysis.** In principle, patterns of spike timing might carry information about upcoming behavior in two different (and non-exclusive) ways. First, information might be encoded by a burst’s onset time relative to the song syllable. Alternately, information might be encoded by the temporal patterning of action potentials within the burst (i.e., the sequence of inter-spike intervals). The direct method outlined above, however, does not differentiate between these two types of temporal encoding. We therefore performed an alternate analysis that isolated the information encoded by patterns of inter-spike intervals, removing the contribution of the burst onset time.

To perform this alternate analysis, we first aligned the spike trains by their first spike, so that spike trains with the same sequence of inter-spike intervals would be registered in time and thus be represented by the same word. (All trials in which zero spikes occurred were assigned the same word.) We then discretized the spike trains at  $dt = 1$  ms and calculated the mutual information between the resulting neural words and behavior as described in the primary analysis above. Across the four acoustic groupings, information estimates of 71% of conditions were unbiased or corrected for bias and 26% of conditions were biased. 2% of conditions failed to produce estimates because there was a lack of coincidences in the response distributions.

### Peak Information from Metric-Space Method

As discussed above, the metric-space and direct methods of computing mutual information differ in their underlying assumptions about the statistical structure of the neural code, and the metric-space method can only produce lower bounds on the signal-response mutual information. Therefore, comparing the values of information computed by the two methods is prone to various problems of interpretation. It is nevertheless instructive to ask whether the direct method estimates greater mutual information than the metric-space analysis, and thus if patterns of multiple spikes carry additional information beyond that in spike pairs, which is discoverable by the metric-space method. To answer this, we calculated the peak metric space information  $\bar{I}_{max}$  which is the mean of  $I_{max}$  across all cases. This is the upper bound on the information detectable through the metric-space method, as the information is maximized for each case independently, rather than finding a single optimal  $q$  for all cases.

### Supporting Information

**Figure S1 Metric-space analysis with normalized and un-normalized distances (alternate analyses).** (a) Shows the same analysis as Figure 2d in the main text, which uses a normalized version of the distance metric originally described by Victor and Purpura. As described in the main text, in this analysis the distance between two spike trains is normalized by dividing by the total number of spikes in both trains (see Materials and

Methods) and demonstrates a significantly higher proportion of temporal cases (filled bars) than rate cases (empty bars; see Results). (b) Very similar results were obtained from an alternate analysis in which the un-normalized Victor-Purpura distance was used.  
(EPS)

**Figure S2 Direct calculation of information reveals greater information at finer temporal resolution (alternate analyses).** In the analysis shown in Figure 3, we combined single- and multiunit data to compute mutual information at different temporal resolutions using the direct method of calculating information. Furthermore, in a minority of cases we partitioned the response distribution to correct for sample size-dependent biases in the entropy estimate derived from the NSB algorithm (see Materials and Methods). In Figure S2, we demonstrate that the key result shown in Figure 3b—that mutual information rises dramatically as temporal resolution increases (Figure S2a)—is qualitatively identical when only single unit data are considered (Figure S2b), when only multiunit data are considered (Figure S2c), or when the datasets that exhibited sample size-dependent biases that could not be corrected are excluded (Figure S2d). The standard deviation of single or multiunit cases alone is, as expected, larger than that of combined results.  
(EPS)

**Data S1 Numerical data underlying main and support-figure figures.** The numerical data associated with all figures except

Figure 4 are provided in .csv files. Information about Figure 4 is provided below. For Figure 4a, the number of times the words are seen in blue group 2 and red group 1, respectively, from top to bottom are 3, 0; 2, 0; 2, 0; 2, 1; 0, 1; 1, 2; 0, 1; 0, 2. For Figure 4b, these counts are 3, 0; 3, 1; 3, 0; 2, 0; 0, 1; 0, 1; 0, 2; 0, 2.  
(ZIP)

**Data S2 Spike timings and acoustic measurements for all cases.** All spike timing and acoustic data analyzed in this manuscript are provided in .csv files, which each contain data from one single-unit or multiunit case. Rows represent the multiple renditions of a song syllable. The first three columns contain the acoustic parameters of pitch, amplitude, and spectral entropy, quantified for each rendition. The following columns present spike timings, which are referenced to the beginning of the premotor window.  
(ZIP)

## Acknowledgments

We thank Michael Long and Robert Liu for helpful discussions and Harshila Ballal for animal care.

## Author Contributions

The author(s) have made the following declarations about their contributions: Conceived and designed the experiments: CT IN SJS. Performed the experiments: SJS DC. Analyzed the data: CT KS IN SJS. Wrote the paper: CT IN SJS.

## References

- Arabzadeh E, Panzeri S, Diamond ME (2006) Deciphering the spike train of a sensory neuron: counts and temporal patterns in the rat whisker pathway. *J Neurosci* 26: 9216–9226.
- Berry MJ, Warland DK, Meister M (1997) The structure and precision of retinal spike trains. *Proc Natl Acad Sci U S A* 94: 5411–5416.
- Borst A, Theunissen FE (1999) Information theory and neural coding. *Nat Neurosci* 2: 947–957.
- Fairhall A, Shea-Brown E, Barreiro A (2012) Information theoretic approaches to understanding circuit function. *Curr Opin Neurobiol* 22: 653–659.
- Lawhern V, Nikonov AA, Wu W, Contreras RJ (2011) Spike rate and spike timing contributions to coding taste quality information in rat periphery. *Front Integr Neurosci* 5: 18.
- Liu RC, Tzovec S, Rebrink S, Miller KD (2001) Variability and information in a neural code of the cat lateral geniculate nucleus. *J Neurophysiol* 86: 2789–2806.
- Mackevicius EL, Best MD, Saal HP, Bensmaia SJ (2012) Millisecond Precision Spike Timing Shapes Tactile Perception. *J Neurosci* 32: 15309–15317.
- Nemenman I, Lewen GD, Bialek W, van Steveninck RRD (2008) Neural coding of natural stimuli: Information at sub-millisecond resolution. *PLoS Comput Biol* 4.
- Reinagel P, Reid RC (2000) Temporal coding of visual information in the thalamus. *J Neurosci* 20: 5392–5400.
- Strong SP, Koberle R, de Ruyter van Steveninck RR, Bialek W (1998) Entropy and Information in Neural Spike Trains. *Phys Rev Lett* 80: 197–200.
- Wang L, Narayan R, Graña G, Shamir M, Sen K (2007) Cortical discrimination of complex natural stimuli: can single neurons match behavior? *J Neurosci* 27: 582–589.
- Georgopoulos AP, Schwartz AB, Kettner RE (1986) Neuronal population coding of movement direction. *Science* 233: 1416–1419.
- Paninski L, Fellows MR, Hatsopoulos NG, Donoghue JP (2004) Spatiotemporal tuning of motor cortical neurons for hand position and velocity. *J Neurophysiol* 91: 515–532.
- Sober SJ, Wohlgenuth MJ, Brainard MS (2008) Central contributions to acoustic variation in birdsong. *J Neurosci* 28: 10370–10379.
- Leonardo A, Fee MS (2005) Ensemble coding of vocal control in birdsong. *J Neurosci* 25: 652–661.
- 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.
- Yu AC, Margoliash D (1996) Temporal hierarchical control of singing in birds. *Science* 273: 1871–1875.
- Chi Z, Margoliash D (2001) Temporal precision and temporal drift in brain and behavior of zebra finch song. *Neuron* 32: 899–910.
- Tumer EC, Brainard MS (2007) Performance variability enables adaptive plasticity of ‘crystallized’ adult birdsong. *Nature* 450: 1240–1244.
- Elemans CP, Mead AF, Rome LC, Goller F (2008) Superfast vocal muscles control song production in songbirds. *PLoS ONE* 3: e2581.
- Elemans CP, Spiets IL, Muller UK, Van Leeuwen JL, Goller F (2004) Bird song: superfast muscles control dove’s trill. *Nature* 431: 146.
- Wohlgenuth MJ, Sober SJ, Brainard MS (2010) Linked control of syllable sequence and phonology in birdsong. *J Neurosci* 30: 12936–12949.
- 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-Com Neural* 8: 127–164.
- Chicharro D, Kreuz T, Andrzejak RG (2011) What can spike train distances tell us about the neural code? *J Neurosci Methods* 199: 146–165.
- Nemenman I (2011) Coincidences and estimation of entropies of random variables with large cardinalities. *Entropy* 13: 2013–2023.
- Nemenman I, Shafee F, Bialek W (2002) Entropy and inference, revisited. *Adv Neural In* 14.
- Brainard MS, Doupe AJ (2000) Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature* 404: 762–766.
- Kao MH, Doupe AJ, Brainard MS (2005) Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature* 433: 638–643.
- Oliveczky BP, Andalman AS, Fee MS (2005) Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol* 3: e153.
- Palmer SE, Kao MH, Wright BD, Doupe AJ (2014) Temporal sequences of spikes during practice code for time in a complex motor sequence. *arXiv preprint arXiv:14040655*.
- Oliveczky BP, Otchy TM, Goldberg JH, Aronov D, Fee MS (2011) Changes in the neural control of a complex motor sequence during learning. *J Neurophysiol* 106: 386–397.
- Schneidman E, Puchalla JL, Segev R, Harris RA, Bialek W, et al. (2011) Synergy from silence in a combinatorial neural code. *J Neurosci* 31: 15732–15741.
- Machens CK, Prinz P, Stemmler MB, Ronacher B, Herz AV (2001) Discrimination of behaviorally relevant signals by auditory receptor neurons. *Neurocomputing* 38–40: 263–268.
- Narayan R, Graña G, Sen K (2006) Distinct time scales in cortical discrimination of natural sounds in songbirds. *J Neurophysiol* 96: 252–258.
- Huetz C, Del Negro C, Lebas N, Tarroux P, Edeline JM (2006) Contribution of spike timing to the information transmitted by HVC neurons. *Eur J Neurosci* 24: 1091–1108.
- Sakata JT, Brainard MS (2006) Real-time contributions of auditory feedback to avian vocal motor control. *J Neurosci* 26: 9619–9628.
- Sakata JT, Brainard MS (2008) Online contributions of auditory feedback to neural activity in avian song control circuitry. *J Neurosci* 28: 11378–11390.



39. Fiete IR, Senn W, Wang CZ, Hahnloser RH (2010) Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. *Neuron* 65: 563–576.
40. Hoffmann LA, Sober SJ (2014) Vocal generalization depends on gesture identity and sequence. *J Neurosci* 34: 5564–5574.
41. Sober SJ, Brainard MS (2012) Vocal learning is constrained by the statistics of sensorimotor experience. *Proc Natl Acad Sci U S A* 109: 21099–21103.
42. Thorpe S, Delorme A, Van Rullen R (2001) Spike-based strategies for rapid processing. *Neural Networks* 14: 715–725.
43. de Ruyter van Steveninck R, Borst A, Bialek W (2001) Real-time encoding of motion: answerable questions and questionable answers from the fly's visual system. Zanker JM, Zeil J, editors. *Motion Vision*: Springer.
44. Brezina V, Orekhova IV, Weiss KR (2000) The neuromuscular transform: the dynamic, nonlinear link between motor neuron firing patterns and muscle contraction in rhythmic behaviors. *J Neurophysiol* 83: 207–231.
45. Burke RE (1981) Motor units: anatomy, physiology, and functional organization. Brooks VB, editor. *Handbook of physiology, the nervous system, motor control II*. Bethesda (Maryland): American Physiological Society.
46. Garland SJ, Griffin L (1999) Motor unit double discharges: statistical anomaly or functional entity? *Can J Appl Physiol* 24: 113–130.
47. Zhurov Y, Brezina V (2006) Variability of motor neuron spike timing maintains and shapes contractions of the accessory radula closer muscle of *Aplysia*. *J Neurosci* 26: 7056–7070.
48. Fee MS (2002) Measurement of the linear and nonlinear mechanical properties of the oscine syrinx: implications for function. *J Comp Physiol A* 188: 829–839.
49. Fee MS, Shraiman B, Pesaran B, Mitra PP (1998) The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature* 395: 67–71.
50. Nemenman I, Bialek W, de Ruyter van Steveninck R (2004) Entropy and information in neural spike trains: Progress on the sampling problem. *Physical Review E* 69: 056111.
51. Spiro JE, Dalva MB, Mooney R (1999) Long-range inhibition within the zebra finch song nucleus RA can coordinate the firing of multiple projection neurons. *J Neurophysiol* 81: 3007–3020.
52. Ashmore RC, Wild JM, Schmidt MF (2005) Brainstem and forebrain contributions to the generation of learned motor behaviors for song. *J Neurosci* 25: 8543–8554.
53. Vu E, Mazurek M, Kuo Y (1994) Identification of a forebrain motor programming network for the learned song of zebra finches. *J Neurosci* 14: 6924–6934.
54. Fee MS, Kozhevnikov AA, Hahnloser RH (2004) Neural mechanisms of vocal sequence generation in the songbird. *Ann N Y Acad Sci* 1016: 153–170.
55. Victor JD (2005) Spike train metrics. *Curr Opin Neurobiol* 15: 585–592.
56. Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W (1996) *Spikes: exploring the neural code*. Cambridge (Massachusetts): MIT Press.
57. Panzeri S, Senatore R, Montemurro MA, Petersen RS (2007) Correcting for the sampling bias problem in spike train information measures. *J Neurophysiol* 98: 1064–1072.
58. Paninski L (2003) Estimation of entropy and mutual information. *Neural Comput* 15: 1191–1253.