

# Temporal Precision and Temporal Drift in Brain and Behavior of Zebra Finch Song

Zhiyi Chi<sup>1,3</sup> and Daniel Margoliash<sup>2</sup>

<sup>1</sup>Department of Statistics

<sup>2</sup>Department of Organismal Biology and Anatomy

University of Chicago

Chicago, Illinois 60637

## Summary

In the zebra finch forebrain nucleus *robustus archistriatalis* (RA), neurons burst during singing. We showed that the internal structure of spike bursts was regulated with a precision of circa 0.2 ms, and yielded alignment of acoustic features of song with a precision of circa 1 ms. In addition, interburst intervals and corresponding syllable durations displayed systematic variation within song (average elongation 0.3 ms/song), and slower “drift” across songs. Systematic variation on even a coarser time scale might be difficult to detect in other systems, but could affect the analysis of temporal patterning. The close relationship between precise timing of individual spikes and stereotypic behavior suggests that song is represented in RA by a temporal code.

## Introduction

Neurons can transmit information in their average rate of firing or in the relative timing of trains of spikes from populations of neurons. The latter, temporal coding hypothesis has been verified in a few systems in the sense that spike timing has been directly related to behavioral decisions animals make (e.g., Carr et al., 1986). It remains unclear to what degree temporal coding is generally present, especially in cortical representations in relation to behavior (e.g., Vaadia et al., 1995; Laurent et al., 1996; deCharms and Merzenich, 1996; Hatsopoulos et al., 1998; Keat et al., 2001). A vigorous discussion has focused on the definitions, scope, and evidence for temporal coding (Singer and Gray, 1995; Usrey and Reid, 1999; deCharms and Zador, 2000) and the statistical analysis of coincident spike events (e.g., Pauluis and Baker, 2000). Temporal coding is thought to potentially contribute to the neural representations of a broad range of behaviors. The choice of which behavior to study, however, can greatly influence the ease with which such codes are detected.

Song in birds is well suited to the analysis of neuronal representations that might incorporate temporal codes. Song is a complex behavior that unfolds in a stereotypic fashion over time, and neuronal dynamics are easiest to understand in the context of corresponding behavioral dynamics. Song is episodic and often repeated, which should facilitate the computationally intensive search for repeated neuronal patterns of activity. In birds such as zebra finches, song is part of courtship behavior and is a particularly intense, controlled behavior (Sossinka

and Böhner, 1980). Precision in behavior should be reflected in precision in neuronal control elements. Finally, acoustic recordings under controlled conditions allow song to be described with high fidelity, which greatly facilitates the description of the behavior.

The descending projections of the nucleus *robustus archistriatalis* (RA) are the forebrain output of the bird-song system, including projections to midbrain and brainstem vocal and respiratory structures, as well as feedback to the thalamus (Wild, 1993). During singing, individual RA neurons exhibit bursts of activity that vary with the identity of the immediately following note (the smallest unit of vocalization, a part of a syllable). The spike-by-spike burst structure of an RA neuron shows great repeatability for multiple occurrences of the same note type, and reliably changes for each different note type that the unit participates in. Thus, with considerable reliability, behavior can be “read off” by examining individual traces of an RA unit’s activity patterns (Yu and Margoliash, 1996).

A detailed quantitative analysis of RA activity patterns during singing has yet to be accomplished, however. Thus, the goal of our analysis was to relate the timing of spike bursts with the timing of acoustic features of song. The analysis of temporal coding generally involves development of appropriate statistical measures that may vary from system to system. In RA, the timing and structure of spike bursts during singing is remarkably precise, and we developed near-optimal solutions for spike alignment. This allowed us analysis of variation in patterns of neural activity and associated behavior that would otherwise have been difficult to detect. We find sub-millisecond precision in RA neuron spike bursts, at least millisecond precision in associated behaviors (timing of acoustic features of syllables), and systematic short-term changes within-song and longer-term drift in these measures.

## Results

Alignment of spike bursts based on burst structure is sensitive to missing spikes. Since we wanted to search for small variation in spike and acoustic timing, we limited analysis to only those sites where excellent isolation was achieved, to minimize missing spikes or variation in estimating spike timing. We focused on the most complex syllables within motifs (repeated sequences of syllables within a song). These long syllables (201–357 ms) were often associated with several distinct bursts of RA activity, and had reliable fast frequency modulations that permitted locating acoustic features with great accuracy. With these constraints, the data derive from 13 RA units, 5 syllable types, 4 adults, collected in two previous studies (Yu and Margoliash, 1996; Dave and Margoliash, 2000). Each site represents 30–120 min of singing.

### Temporal Precision of RA Pre-motor Activity and Associated Singing

When zebra finches sang, RA neurons reliably exhibited discrete groupings of high-frequency spiking activity,

<sup>3</sup>Correspondence: chi@galton.uchicago.edu

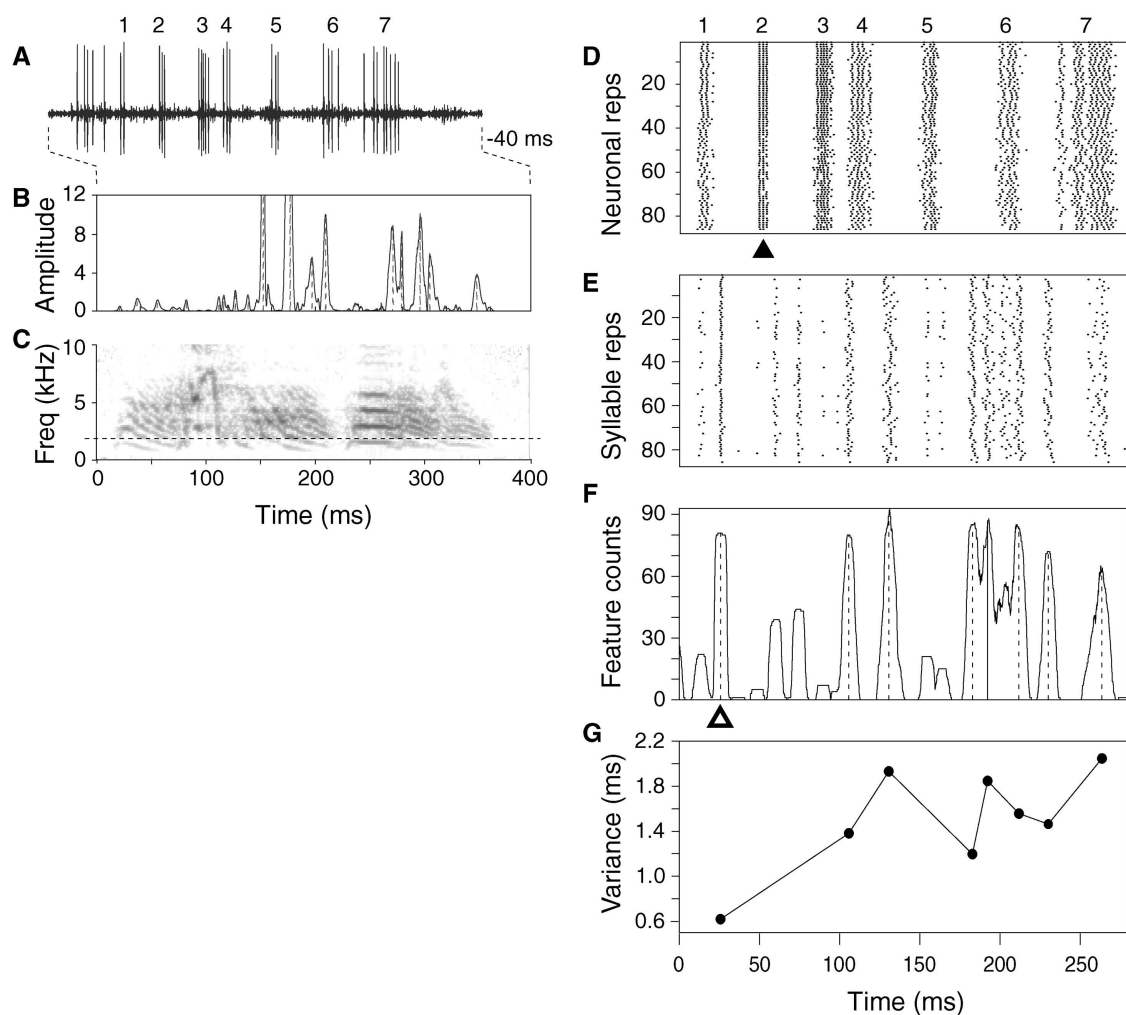


Figure 1. Aligning Acoustic Features by Neuronal Activity

Right column, finding acoustic features. (A) Raw trace of neuronal activity of unit 1, and is offset relative to the other panels by 40 ms to roughly compensate for pre-motor lead time. The small numbers identify the corresponding bursts in panel (D). (B) Amplitude profile in a band of  $1641 \pm 78$  Hz of the corresponding syllable D that the bird sang. The vertical dashed lines indicate points where the signal amplitude exceeded a threshold, indicating an acoustic feature. (C) Spectrograph of the syllable. Dashed line indicates 1641 Hz. Left column, aligning spikes and acoustics. (D) Spike activity during 83 repetitions of syllable D for unit 1. Spikes in the second burst (filled arrowhead) have been aligned (hence the near-perfectly vertical lines) and all other spikes in each repetition have been shifted in time according to the alignment for that repetition. (E) Local maxima were identified in the amplitude profile at  $1641 \pm 78$  Hz of syllable D. For each point that is plotted, the local maxima exceeded a fixed threshold. These points, a subset of all local maxima, are termed acoustic features. The acoustic features in syllable D have been aligned relative to the alignment of the second spike burst. (F) The number of acoustic features detected within a small window as the window is swept across each syllable D that the bird sang. The vertical dotted lines represent the timing of local acoustic features, i.e., features that exceed a threshold count of 70% (see Experimental Procedures). (G) The variance in the timing of the eight local acoustic features represented in (C). Note that the first local acoustic feature (open arrowhead in [C]), which is closest in time to the aligned spike burst, has the lowest variance.

separated by much longer intervals of complete silence. Each group of spikes is called a burst. RA neurons typically exhibited multiple bursts during each complex syllable. Each burst typically contained many spikes, and had a unique internal structure (sequence of interspike intervals); thus, each burst had a unique identity and could be classified as belonging to one of a discrete set of burst types. By matching the acoustic and associated neuronal data, a one-to-one correspondence was established between syllable type and sequence of spike burst types (Yu and Margoliash, 1996).

The dataset of songs and associated RA spike times

contained labels identifying each syllable as to type, and onset and offset times, assigned by a human expert. To examine the structure of spike burst timing, we extracted all the exemplars of a given syllable type and the associated spike data. In prior studies, syllable onsets were adjusted based on maxima in cross-correlograms of spectrographs between all syllables of the same type. When the spike trains associated with the syllables were adjusted accordingly, an internal structure to the bursts could be detected. Furthermore, when the onset timing of spike bursts were aligned based solely on the timing of the spikes (i.e., ignoring the timing

of the acoustics), this revealed fine temporal patterns within the bursts that were otherwise difficult to detect (Yu and Margoliash, 1996; Dave and Margoliash, 2000). The previous procedures aligned spike bursts so as to minimize the temporal distance to a single “referent” burst. We developed an optimization procedure that did not require a referent burst (see Experimental Procedures). Applying this procedure to all 59 bursts for all 14 syllable types (13 units—two complex syllable types were available for the single unit recorded from the bird RA01), across all bursts the average variance in the timing of the first spike within each burst was remarkably small, only  $0.23 \pm 0.17$  ms. The optimization procedure was unbiased in the sense that it put equal weight on all spikes. This demonstrates that the internal structure of spike bursts is highly regulated, although there may be a slight tendency for increased variance of timing for the latter spikes of bursts (e.g., Figures 1D, 2A, 2B, 3A, and 3B).

The small variation in spike timing revealed with the optimization procedure suggested that there might also be similarly small variation in the acoustics of song. Previously, cross-correlation of spectrographs of notes had failed to reveal such variation (Yu and Margoliash, 1996), but this failure could have been due to insufficient temporal resolution. Thus, a new procedure was developed to identify with high temporal precision the timing of individual acoustic features of song. Using this procedure, across all exemplars of a syllable type, the absolute time of each acoustic feature of song was adjusted by the shift in time required to align the corresponding spike burst. This yielded a fine temporal pattern in the acoustic features. For example, for unit 1, when the bird RA01 sang syllable D multiple times, the same acoustic features could be detected for each exemplar of syllable D. When the timing of acoustic features was represented by a raster plot, parallel lines representing acoustic features were observed, resembling the parallel lines representing spike timing (Figure 1E). A similar phenomenon was observed for all units.

The alignment of acoustic features based on spike burst timing strongly depended on choosing a spike burst that was temporally close to the acoustic feature. For example, for unit 13 (bird YU49), aligning the second spike burst, which occurred toward the beginning of syllable C (Figure 2A), resulted in good alignment of initial acoustic features of syllable C and some disorder in terminal acoustic features of syllable C (Figure 2C). Conversely, aligning the third spike burst, which occurred toward the end of syllable C (Figure 2B), resulted in good alignment of terminal acoustic features of syllable C and some disorder in initial acoustic features of syllable C (Figure 2D). Quantitatively, the variance in the timing of the aligned acoustic features was lowest for those features that were closely temporally associated with the aligned spike burst. The variance increased for acoustic features that were temporally more distant from the aligned spike burst (e.g., Figure 1G). This effect was seen for all 12 syllables (one syllable had only a single burst), so that positive correlations were observed for each syllable when comparing the temporal distance between all acoustic features and each of the aligned bursts with the “acoustic scatter” (defined as the standard error of the timing) of the features. For 8/12 syllables,

the correlations ( $r = 0.32 \pm 0.138$ ) were significant ( $p < 0.05$ ).

The tight coupling between acoustic features and neuronal activity permitted analysis of the precision of the acoustics (behavior) using the superior temporal resolution of the neuronal data as the standard of measurement. We defined a “reliable” local acoustic feature of a syllable type as one that was detected in at least 70% of the exemplars of that syllable type (this eliminated cases of excessive contamination from female calling and other noise). We chose small bin widths to distinguish temporally adjacent acoustic features (see Experimental Procedures, Figure 1F). We considered only cases where a reliable local acoustic feature occurred within  $\pm 50$  ms of the onset of a spike burst, after the adjustment for 40 ms pre-motor lead. On this basis, across all neurons, adjusting the timing of the acoustic features based on the alignment of the associated spike bursts resulted in a variance in the acoustic features of only  $1.04 \pm 0.28$  ms, a remarkable precision of vocal behavior (Table 1). For each neuron, the acoustic scatter was less than 1/4 of the bin width, hence a 95% confidence interval of the timing of acoustic features relative to the timing of neuronal bursts fell entirely within the bin. Collectively, these observations demonstrate for the first time quantitatively spike timing of RA neurons that is closely associated with local acoustic features of a syllable type. Such temporal precision will generally hold for syllables of the motifs of songs but not for introductory notes of song (Yu and Margoliash, 1996).

### Temporal Drift in RA Activity

Theories of temporal coding and pattern-finding algorithms often assume stationarity of spike patterns. The precision with which we were able to measure neural activity allowed us to directly test this hypothesis. Given such tight regulation of intraburst spike timing and associated acoustics, it was surprising that when RA neurons exhibited multiple bursts within syllables, we often observed a systematic variation in the onset timing of adjacent bursts. In raster plots of spike trains, the variation in onset timing manifested itself as an overall change in intervals between burst onsets. For example, in Figure 3, for each of two syllable types the first spike bursts are aligned, maximizing the observed structure in those bursts. All other spikes are plotted relative to the first bursts without further adjustment. There is a clear pattern in the timing of the other “target” bursts relative to the aligned first bursts, so that each time the bird sang the syllable, the onsets of the target spike bursts “drifted” away from the onset of the aligned burst. The sense of the drift (positive or negative) depended strictly on the position of the target burst relative to the aligned burst. Thus, in all cases where drift was observed, the direction of the drift in target bursts was reversed when alignments were computed on bursts preceding or following target bursts. The overall direction of drift was typically toward increasing interburst intervals.

To assess the statistical significance of the apparent drift in burst timing, we calculated the time intervals between bursts, choosing a well-aligned single spike within a burst and using the time of that spike as the time of the burst. Typically, the onset spike was chosen.

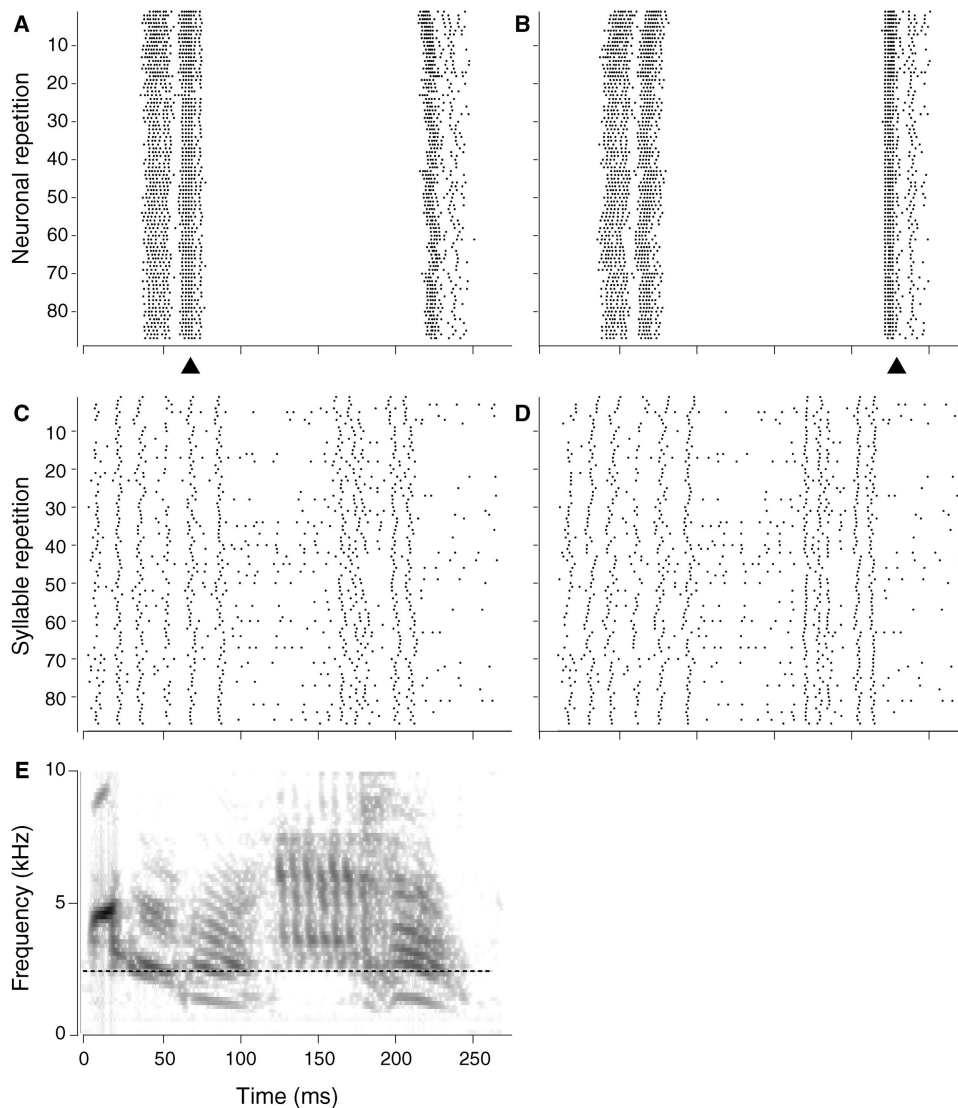


Figure 2. Neural Activity and Acoustics Associated with 86 Repetitions of Syllable C of Unit 13

In (A), neural activity has been aligned to the second burst (arrowhead) and in (C), local acoustic features are aligned relative to (A). In (B), neural activity has been aligned to the third burst (arrowhead) and in (D), local acoustic features are aligned relative to (B). The noise in (C) and (D) results from false acoustic features detected in female calling. Note that the best aligned local acoustic features in (C) and (D) are those closest to their respective aligned spike bursts. In (E), a spectrograph of syllable C; the dashed line indicates 2578 Hz. In this case, the relatively high frequency chosen for analysis captures more acoustic features.

In most cases, the first and last spike bursts associated with a syllable were chosen for analysis, to obtain the largest observable drift. The drift was analyzed using the standard linear regression of the duration of the time interval between bursts on the repetition number of the syllable. By this analysis, for 11/13 neurons (all four birds), the spike data exhibited significant drift (Table 2). The magnitude of the drift was  $0.107 \pm 0.215$  ms/repetition of an average burst interval of 154 ms, or  $0.695 \pm 1.396$  ms/s of song. This number, however, averages out a larger variation in relative timing observed when analyzing zebra finch songs over shorter intervals of time (see below), and is inflated by an outlier unit (no. 5) with unusually large drift. Unit 10 was excluded from this analysis because it had only a single

burst during the complex syllable. Two units (no. 4, 9) failed to exhibit within-syllable drift. For unit 4, however, the bird sang only briefly while unit isolation was maintained (16 repetitions of its complex syllable), hence the power of the statistical test was reduced. For a fourth unit (no. 7), drift was readily apparent throughout most of the recording; however, there was a sudden discontinuity (a “re-setting”) that occurred near the end of the recording. The drift for this unit was highly significant when considering the recording prior to the re-setting, but not when considering the entire data set for this neuron (Table 2). Such re-setting was also observed in other neurons (see below). Thus, only one unit (no. 9) presented sufficient data where we would expect to detect drift but failed to do so.

Table 1. Temporal Alignment of Acoustic Features

Bird	Unit/Syllable	Number of Acoustic Features Aligned	Mean Scatter (ms)	SD of Scatter
RA01	1/D	69	1.01	0.25
	1/E	62	1.01	0.26
RA02	2/D	111	0.93	0.28
	3/D	123	1.13	0.21
	4/D	171	1.13	0.23
	5/D	38	0.99	0.32
RA16	6/FG	61	0.79	0.32
	7/FG	17	1.16	0.34
	8/FG	41	1.20	0.32
	9/FG	71	0.99	0.39
	10/FG <sup>a</sup>	0	NA	NA
YU49	11/C	24	1.07	0.20
	12/C	26	1.07	0.22
	13/C <sup>b</sup>	NA	NA	NA

<sup>a</sup>All bursts occurred at the end of the syllable; no local acoustic feature occurred within  $\pm 50$  ms.

<sup>b</sup>Too much contamination from female calls to define local acoustic features.

Drift between spike bursts was observed not only within syllables but in bursts associated with different syllables, throughout the motifs each bird sang. For example, in Figure 4, longer sections of song with multiple bursts representing neural activity during entire motifs are shown for three of the exceptional neurons described above (units 4, 9, 10), for which within-syllable drift was not observed. Because in this example the time interval plotted is longer, the drift and re-setting of spike activity is more obvious. Unit 4 had few repetitions, but time axis compression was apparent when considering activity across the entire motif. Unit 9 failed to exhibit drift for its complex syllable, but relatively large changes in spike timing starting at about repetition 30 were apparent when the timing between syllables was considered. Unit 10 had only a single burst for each of three syllables, but it clearly exhibited time axis expansion when timing between syllables was considered.

To quantify the drift across syllables, we compared the timing of the last spike burst in the complex syllable with the first spike burst in a preceding syllable of the same motif. (Some units did not burst in the syllable immediately preceding the complex syllable.) In addition, for two units, the bird sang the complex syllable in two distinct motifs, which increased the number of cases. This resulted in 16 cases (13 units), of which 15 showed significant intersyllable drift ( $p < 0.032$ ). The magnitude of the drift was  $0.18 \pm 0.167$  ms/repetition for a 364 ms average interval, or  $0.495 \pm 0.460$  ms/s of song. Thus, intersyllable drift was confirmed for each unit bar one, emphasizing that time axis variation was observed throughout motifs.

The analyses suggested that drift was an ongoing, cumulative phenomenon that occurred at a basal rate throughout song under the conditions of these recordings. As one test of this hypothesis, we assessed whether interburst intervals covaried. Indeed, for each site that exhibited statistically significant drift, interburst intervals exhibited a linear relation in their duration. For example, for the activity of unit 2 during syllable D, when the spike data were aligned by the third burst in syllable D, the timing of the other bursts associated with syllable D drifted away from the third burst. A scatter diagram

relating burst intervals 1–3 and 3–4 graphically demonstrates this relationship, which had a correlation coefficient of 0.40 (Figure 5).

To quantitatively test the hypothesis that interburst intervals covaried, we compared pairs of nonoverlapping burst intervals, eliminating artifactual inflation of the correlation coefficients that would have otherwise resulted from the overlapping interval. The intervals were defined by three bursts. The first and the third bursts were chosen as bursts for which significant relative drift was established. Where possible, the intervening second bursts were chosen to be about equidistant from the other two. Under these criteria, high correlation coefficients were commonly observed comparing burst intervals within complex syllables. The absolute value of the correlation coefficient was  $0.41 \pm 0.21$  for 10 pairs of intervals for 10 units (three birds). (Four syllables were excluded from this analysis: for unit 1 (two syllables), the drift between the intervening burst and the extreme bursts was not significant; for unit 4, there were insufficient repetitions; for unit 10, there was only a single burst.) Interestingly, for several cases, some pairs of burst intervals were negatively correlated, even while the total duration of the intervals exhibited positive drift. For example, during syllables FG, unit 7 generated three bursts. The interval between the first and the last burst had highly significant drift ( $p < 0.0001$ ). The correlation between the two nonoverlapping intervals was  $-0.54$ . This implies that during drift, compression and expansion of the time axis occurred concurrently. Furthermore, the positive correlation ( $\alpha$ ) we observed between nonoverlapping intervals ( $X$  and  $Y$ ) implies that the correlation between  $X$  and  $X + Y$  is  $1 + \alpha$ . This implies that longer burst intervals are associated with larger magnitudes of neuronal drift.

#### Drift Occurs over at Least Two Time Scales

The preceding data demonstrate changes in intervals between spike bursts within syllables. These changes could result from dependencies on elapsed time, the temporal dynamics of singing, or a combination of the two. To assess these possibilities, we examined the timing of spike bursts relative to elapsed time (“wall

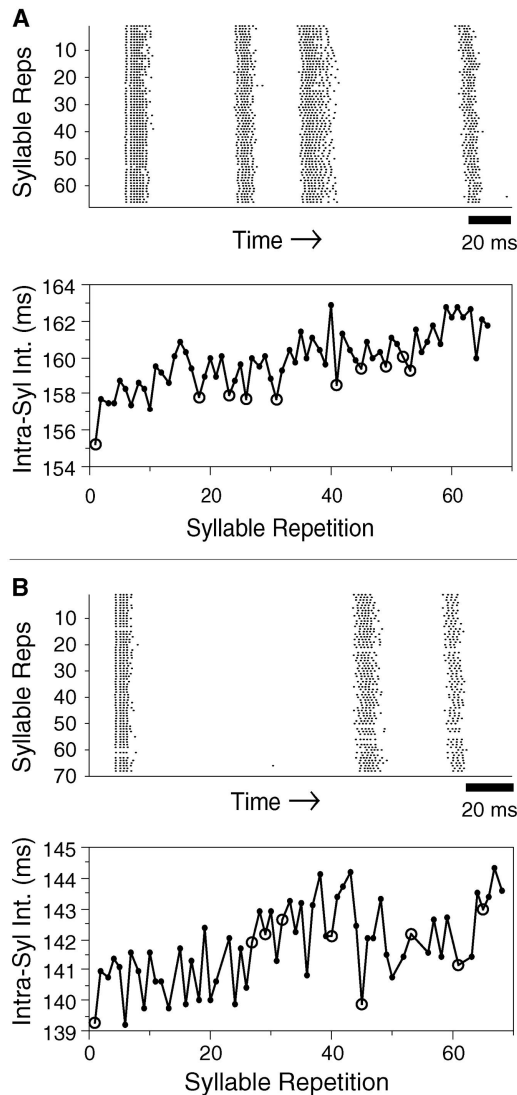


Figure 3. Neuronal Drift during Syllables

Neural activity for unit 2 during 66 repetitions of syllable D (bird RA02) are shown in (A), and for unit 8 during 68 repetitions of syllable FG (bird RA16) are shown in (B). The raster plots show the timing of 4 (A) or 3 (B) bursts of neuronal activity that occurred during the corresponding syllables; each raster is aligned on the first burst. The line plots show the time interval between the first and last burst (defined as the time between the first spike of the first burst and the first spike of the last burst). Note there is a strong tendency for the intrasyllable intervals to increase over the time course of the recordings. Songs separated by silent intervals of  $>30$  s are defined as bouts. The first intrasyllable interval in each bout is marked by a large, open circle. For unit 2, the first intrasyllable interval of a bout is the shortest duration interval of that bout and is shorter than the final intrasyllable interval of the previous bout (i.e., occurs at a local minimum); for unit 8, there is also such a trend, but there are many exceptions as well.

clock" analysis), and relative to syllable positions within song. Wall clock analysis will produce clumps of syllables since much larger intervals separate song bouts than songs within bouts. This results in clumped data in plots of drift versus elapsed time. In the insets for Figure 6, the clumps of data are numbered for the two representative units plotted.

We observed that when birds sang multi-motif songs, the burst intervals associated with the complex syllables of the second motifs were typically larger than the burst intervals associated with the complex syllables of the initial motifs. The changes in the burst intervals of 3rd and subsequent motifs were smaller and not always positive, but the trend within individual songs was toward longer intervals. To show this graphically for two neurons, in Figure 6, subsections of burst intervals versus syllable repetition data are numbered, each number representing a separate bout of singing. It can be seen that burst intervals often increase sharply in syllables following the first motifs of songs. Considering the entire database, there were 200 complex syllables detected in motifs following first motifs of songs. For about 75% (154/200) of syllables, the associated neuronal activity had longer burst intervals, with an increase of  $1.45 \pm 1.81$  ms per syllable. For the 25% of syllables with shorter burst intervals, the decrease was  $-0.71 \pm 1.41$  ms. Overall, the average change in burst interval per syllable in the following motifs was  $0.94 \pm 1.25$  ms, and the distribution was approximately normal. This measure of change is much higher than the estimate of drift (0.11 ms per syllable repetition; see above) that fails to account for the structure of singing. The difference between the two measures can be accounted for by the re-setting of burst intervals at the start of each song.

The change in timing between bouts of songs was more variable. Given that the number of bouts of songs for each unit in the dataset was small, it was difficult to establish any statistically significant pattern for the inter-bout variation. In some cases, there was a re-setting of the burst timing across song bouts, which decreased the duration of a burst interval. For unit 2 (Figure 3A), for example, the re-setting was insufficient to completely offset the increase caused by the change in previous song bouts, leading to a low-magnitude drift of spike burst timing in one direction. It appears that for this neuron, as well as many others, that the activity tended to have different overall drift in different periods. Unit 2 seemed to have a slower overall drift in the latter part of the period of recorded activity. On the other hand, unit 8 had two periods of rapid drift during singing, with a long period absent of singing in between. The first song of the second period had short intrasyllable intervals (Figure 3B, open circle at repetition 45). The re-setting between the bouts of singing almost completely offset the increase in the duration of burst interval that occurred in the first period.

#### Acoustic Drift and Neural Drift

If spike burst timing drifts relative to song acoustics, then the features of song represented by spike bursts would be dynamic over the time scale of changes observed in the neuronal recordings. Alternatively, we hypothesized that during these recording sessions, the timing of the acoustics of song changed in a manner consistent with the neural drift we had observed. We calculated the correlation between changes in timing of the acoustic features and timing of spike bursts. In 10 cases where recordings were relatively uncontaminated with external sounds, a strong correlation was observed ( $0.77 \pm 0.11$ , Table 3). Thus, we demonstrated that

Table 2. Neuronal Drift in Complex Syllables

Unit/Syllable	Drift (ms/rep)	95% Confidence Interval	r	p-Value
1/D	0.01	[0.003, 0.017]	.25	0.008
1/E	0.028	[0.005, 0.050]	.33	0.016
2/D	0.061	[0.048, 0.075]	.76	<0.0001
3/D	0.013	[0.002, 0.023]	.28	0.021
4/D	-0.12	[-0.411, 0.164]	.24	0.371
5/D	0.75	[0.367, 1.135]	.83	0.002
6/FG	-0.027	[-0.052, -0.003]	.35	0.029
7/FG	0.01	[-0.009, 0.028]	.12	0.317
*w/o last 11 reps	0.047	[0.027, 0.067]	.54	<0.0001
8/FG	0.039	[0.025, 0.052]	.59	<0.0001
9/FG	0.014	[-0.009, 0.038]	.20	0.229
10/FG	NA			
11/C	0.074	[0.042, 0.106]	.57	<0.0001
12/C	0.092	[0.045, 0.139]	.54	0.0003
13/C	0.039	[0.024, 0.053]	.50	<0.0001

changes in neuronal timing were reflected in concomitant changes in acoustic timing. For these 10 cases, the average magnitude of acoustic drift (measured between the first and last reliable acoustic features of a syllable) was  $0.057 \pm 0.051$  ms/repetition of an interval of average duration of 140 ms, or  $0.41 \pm 0.36$  ms/s of song. Calculating over the same eight units (nos. 1/D, 1/E, 2, 3, 7, 8, 11, 12), for which significant acoustic drift (0.326 ms/s song) and neuronal drift (0.286 ms/s song) was detected, yields similar values.

## Discussion

In nucleus RA of the bird song system, the fine temporal pattern of spike bursts are associated with specific features of behavioral output. Different spike patterns are observed with different microscopic acoustic features of songs, probably relating to different notes within syllables (Yu and Margoliash, 1996). We used the timing of all spikes within a burst to uncover the remarkable precision of both spiking activity and behavioral output. This result would not have been obtained if intervals within bursts were shuffled, as in a test for rate coding. Although part of the distinction between rate and temporal coding centers on the saliency of windows of integration of 10–20 ms (rate coding) versus 1–2 ms (temporal coding) (see deCharms and Zador, 2000), the standard error of neuronal timing we have observed is roughly an order of magnitude smaller ( $\sim 0.2$  ms). A 10 ms bin would completely encompass entire RA bursts in many cases. These data support the conclusion that in zebra finches, singing behavior is described in part by a temporal code observed at the level of single RA neurons. The data do not describe the form or mechanisms of the code.

## Neural Coding of Stereotypic Behavior

To achieve highly stereotypic singing performance, the data demonstrate that the nervous system uses much of the physiological range available to it. Although RA neurons have average ongoing discharge of approximately 30 Hz, discharge rates drop to zero just prior to and then throughout a song, punctuated by bursts of spikes. Thus, for the neurons we have analyzed, phasic bursting activity is the only activity pattern by which

behavior is expressed (cf. Shadlen and Newsome, 1998). Bursts have been proposed as a particularly reliable mechanism for transmitting information (Lisman, 1997). Spike bursts in RA can vary from one to at least 13 spikes. We observed that whatever burst was aligned, the variance of timing of nearby acoustic features was reduced. Thus, all bursts—even single spikes—appear to carry information about the timing of acoustic features (cf. Lisman, 1997). The frequency within spike bursts approaches biophysical limits, so that instantaneous spike frequencies can exceed 700 Hz, neurons are driven into their relative refractory periods, and the terminal spikes within bursts can approach spike failure (Yu and Margoliash, 1996). Presumably, both inhibition and excitation are precisely controlled to produce the precision of neuronal activity relative to behavior that we have observed (Spiro et al., 1999).

Beyond these features of activity patterns of individual neurons, the data give some insight in population coding in RA. In monkeys, it has been observed that the independent contribution of individual primary motor cortex neurons summed vectorially predicts the direction of movement (Georgopoulos et al., 1986). This implies that relatively little information regarding movements is encoded by individual motor cortex neurons, and that variation in activity of individual neurons is averaged out by the summation. In contrast, behavior is well represented in the activity of individual RA neurons, so that the precise timing of acoustic features of song is reliably represented in single neuronal traces. The systematic change in the timing of acoustic features can be accounted for by the systematic change in the timing of spike bursts. The RA neurons described here vigorously participated in singing (Yu and Margoliash, 1996; Dave and Margoliash, 2000). (There may be variation in RA neuronal activity patterns that relate to different classes of cells. Neurons in the dorsal cap region of RA that project to the midbrain may have different patterns of activity than the rest of RA; unpublished data. In addition, RA projects to brainstem respiratory centers and activity of some RA neurons may be related to respiratory rhythms.) Ongoing activity of nearby RA neurons results in part from intrinsic oscillations (Mooney, 1992), but activity patterns of RA neurons during singing are unlikely to be indepen-

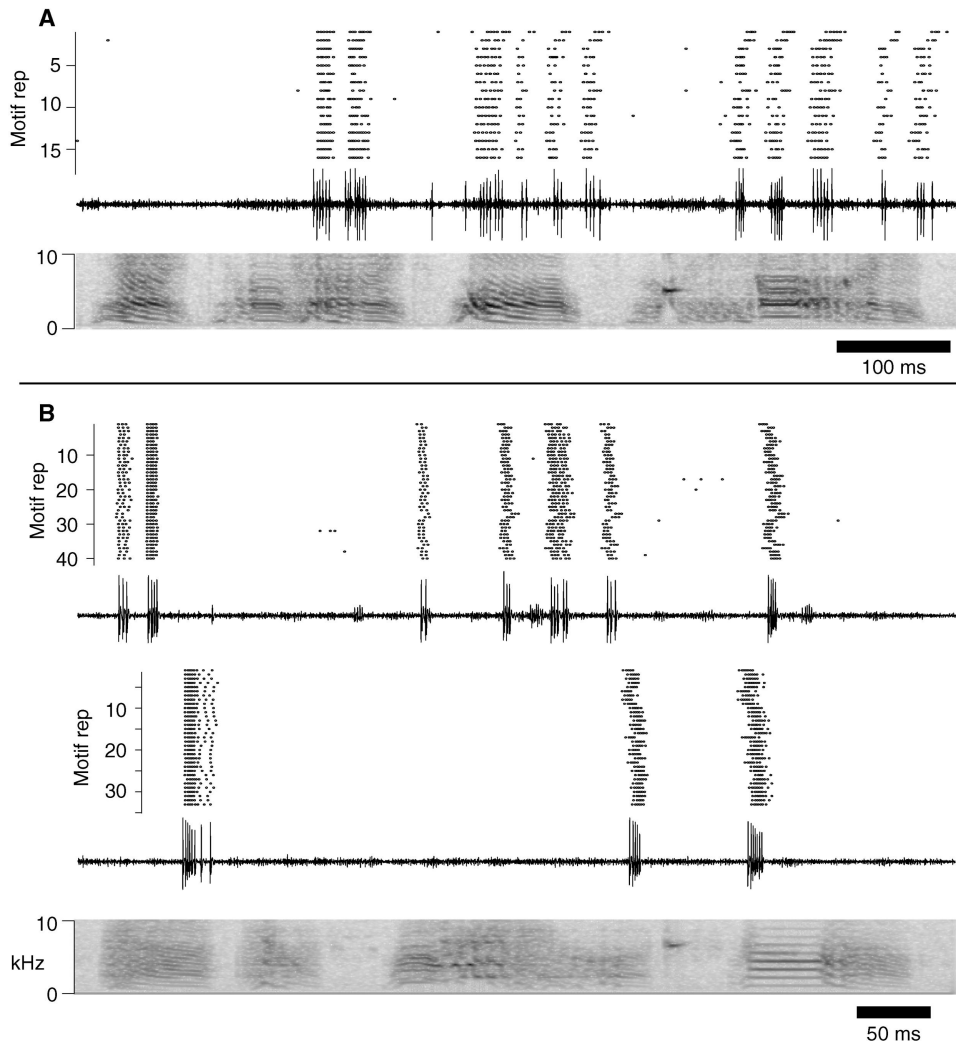


Figure 4. Neuronal Drift during Motifs

Neuronal activity patterns for three cells that failed to show drift within syllables show drift across syllables within motifs. (A) Activity is aligned for unit 4 by the first burst; (B) for unit 9 (top) by the second burst, and for unit 10 (bottom) by the first burst. Although statistically significant drift could not be demonstrated when considering individual syllables, statistically significant drift was demonstrated when considering activity across several syllables. Exemplar motifs are shown in the spectrographs.

dent. Singing is the result of the activity of multiple RA neurons. Tight temporal correlation between neurons can be presumed from the remarkable precision observed in the single cell data, and this has been observed directly in cases where pairs of neurons have been recorded during singing (Dave and Margoliash, 2000; A. Leonardo and M.S. Fee, 2000, Soc. Neurosci., abstract). These correlations are likely to result from groups of neurons participating in local pattern-generating circuits.

These differences between primate and songbird motor areas may result in part from methodological differences. In the primate experiments, the many degrees of freedom of movement were artificially reduced, and precise records of movements were difficult to obtain. In contrast, our data suggest that spike bursts are unitary events, and that the temporal unfolding of behavior is represented by the timing and identity of spike bursting patterns. The limitations of this organizational scheme

for RA should be emphasized, however. Whereas RA spike bursts are highly synchronized throughout the syllables of the motifs of songs, burst patterns are far less structured during the introductory notes of song, as if the system is preparing for singing motifs. This could imply a computational role during note production that is distinct from a representational role during syllable production (deCharms and Zador, 2000). Furthermore, the extreme precision in timing we observed for syllables was obtained when male zebra finches sang their songs “directed” toward females. In the small number of recordings we examined of birds singing “undirected” songs (no females present), RA activity patterns were less reliable (data not shown; see also Hessler and Doupe, 1999). This suggests that extreme temporal precision such as we have observed is maintained only under the most challenging of behaviors. This opens the possibility that temporal coding is present in other



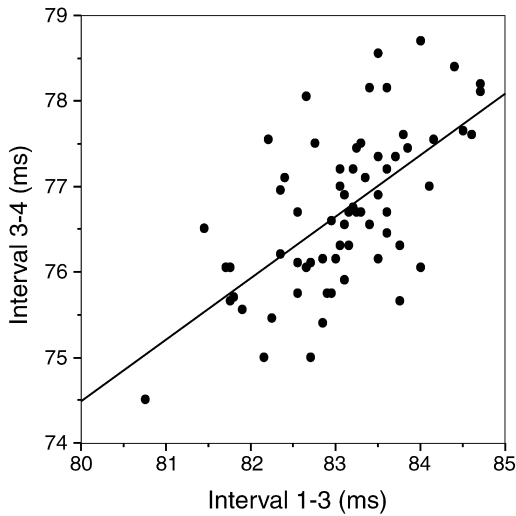


Figure 5. The Duration of Two Nonoverlapping Intervals Are Plotted for Activity of Unit 2 during Syllable D, Demonstrating a Linear Relation. The slope of the regression line was 0.72 and the correlation was highly significant ( $p < 0.0001$ ).

systems but is more difficult to observe (e.g., Hatopoulos et al., 1998).

#### Statistical Analysis of Temporal Precision and the Importance of Behavior

Our results demonstrate that even during directed songs, there are subtle but systematic changes in the relative timing of spike bursts associated with different notes. For bursts that are far apart, the accumulation of the change can be more than a few milliseconds. Failing to identify that the change is systematic could have resulted in its being misinterpreted, as evidence *against* the existence of fine temporal pattern of neuronal activity. By extension, if the neural mechanisms for temporal control of behavior in zebra finches shares similarities with those in other systems, then possibly those systems also experience temporal drift. If such drift is not directly observable, any statistical measure of temporal precision of spiking activity must take into account a hidden drift, presumably on a time scale not finer than that we have reported here.

In most systems, the neuronal activity is much less phasic than we have observed in RA, making it more challenging to find fine temporal pattern in neuronal activity. In such cases, it is particularly important to record precise details of the associated behavior, to facilitate identification and quantification of simple and informative features of behavior, yet this is not always accomplished. A precise description permits direct tests of the temporal precision of behavior associated with neuronal activity, and provides prior knowledge as to the precision expected from the neuronal activity. In our study, identifying the appropriate features of vocalizations was the key to establishing the temporal precision of behavior associated with neuronal activity. The features used were simple (points on the frequency-time plane), rather than high-dimensional vectors more commonly used, yet these simple features contained much

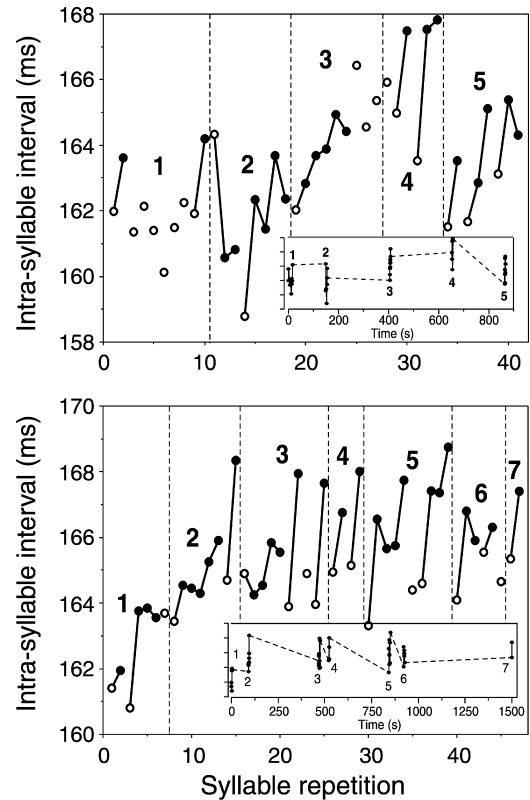


Figure 6. Intrasyllable Intervals Increase during Singing

Data are from unit 12, syllable C (top) and unit 11, syllable C (bottom). Each inset plots the intervals between neuronal bursts during syllable C as a function of the time of occurrence of each syllable C relative to the onset of the recording session. Clusters of syllables are evident, representing bouts of singing. In the main graphs, intervals are plotted relative to the repetition number of syllable C. Each bout is numbered and is separated from the other bouts by vertical dotted lines. Lines connect all syllables within a song. Syllables occurring during the first motifs are represented by open circles and syllables in other motifs are represented by closed circles. Note that in most cases, the intervals between neuronal bursts tend to increase throughout the song.

information regarding the acoustics. The acoustic features were reliably repeated and independent of specific neurons being recorded, facilitating detection by statistical tests.

The temporal patterning we have described for RA neurons might also be present in other systems but be harder to detect. For neuronal activity that is not as structured as in RA, the selection of appropriate statistics and analysis of the power of the statistic is particularly important. In several recent studies, it was concluded that precise repeating patterns of cortical spikes occurred at chance levels; however, the issue of the power of the statistics used was not addressed (Oram et al., 1999; Baker and Lemon, 2000). Most of the statistics used are either the total count of patterns with a certain number of spikes repeating a certain number of times, or the count of a specific nonmaximal pattern. The most powerful statistics, with respect to an alternative hypothesis supporting the existence of fine temporal pattern, are likely to be some “extreme” statistics, i.e., sta-

Table 3. Correlation between Acoustic and Neuronal Relative Timing

Unit/Syllable	$r^a$
1/D	0.75
1/E	0.68
2/D	0.96
3/D	0.84
4/D	0.74
7/FG	0.89
8/FG	0.85
9/FG	0.73
11/C	0.72
12/C	0.61

<sup>a</sup>All values significant at  $p > 0.0009$ .

tistics coming from the tail of the distribution on a certain random variable (Date et al., 1998). For this reason, the count of a pattern that has the maximum number of spikes among all the repeating patterns might be a more powerful statistic.

### Functional Implications of Temporal Drift

The drift in neuronal and acoustic behavior we observed was highly statistically significant, but does it have functional significance? Variation in male song affects female choice (Searcy and Yasukawa, 1996). Zebra finches modulate song duration based on social context. Males sing songs more rapidly and with 10–35 ms shorter motifs when directing those songs to females than at other times (Sossinka and Böhner, 1980). In addition, we have found during directed songs slight but systematic changes in duration within syllables and between syllables. This suggests that the mechanism that controls song duration during undirected singing is more precisely regulated during directed singing.

Male characteristics evolving under sexual selection, such as song, often result in traits with extreme phenotypic characteristics. There is evidence that song dynamics reflect the limits of performance abilities (e.g., Suthers et al., 1999). In zebra finches, the directed song is an intense performance by the male to attract the female. The largest changes in duration we detected within motifs of directed songs were about an order of magnitude smaller than the changes between directed and undirected songs. This indicates that males strive to maintain far greater precision of syllable duration during directed singing than the range of variation that is available to them. Potentially, females are sensitive to the variation of syllable duration within song, which have the largest magnitude, can be directly assessed without reference to a memory of the undirected songs of an individual, and would be easier to detect than the smaller differences spread across songs. Alternatively, males maintain variation in song dynamics below a threshold of detection by females.

### Experimental Procedures

The key to our study was to find the fine temporal structure in the pre-motor activity in RA and the associated singing. To achieve this, methods were developed based on alignment of neuronal data and high temporal resolution analysis of acoustic data. All algorithms were implemented in the MATLAB language (The MathWorks, Natick, MA).

### Analysis of Neural Activity

First, we introduce a distance for the spike trains. Let  $S = \{s_1, \dots, s_n\}$  be a sequence of spike times extracted relative to the start of an arbitrary interval, and  $T = \{t_1, \dots, t_m\}$  be another sequence of spike times extracted relative to the start of another arbitrary interval. The distance from each spike in  $S$  to  $T$  is defined as

$$d(s_i, T) = \min \{|s_i - t_j|, j = 1, \dots, m\}$$

and the distance from  $S$  to  $T$  as

$$D(S, T) = \sum_{i=1}^n d(s_i, T) \quad (1)$$

In general,  $D$  is not symmetric, i.e.,  $D(S, T) \neq D(T, S)$ . Different weights on the terms in the summation,  $D(S, T) = \sum_{i=1}^n w_i d(s_i, T)$ , could yield better alignment if additional knowledge is available on which spikes are more important in information coding. In this study, we only considered uniform weights on spikes. For an arbitrary number of spike trains  $S_1, \dots, S_N$ , define

$$D(S_1, \dots, S_N) = \sum_{i \neq j} D(S_i, S_j) \quad (2)$$

and by alignment we mean the following optimization problem,

$$(t_1, \dots, t_N) = \underset{(\Delta_1, \dots, \Delta_N)}{\operatorname{argmin}} \{D(S_1 + \Delta_1, \dots, S_N + \Delta_N)\} \quad (3)$$

with  $S + \Delta = \{s_1 + \Delta, \dots, s_n + \Delta\}$  for any  $S = \{s_1, \dots, s_n\}$  and number  $\Delta$ . While a global optimal solution for (3) is difficult to find, by observing that

$$D(S_1 + \Delta_1, \dots, S_N + \Delta_N) = \sum_{j=1}^N \sum_{i \neq j} D(S_i + \Delta_i, S_j + \Delta_j)$$

a sub-optimal solution can be found by starting with  $\Delta_n = 0$ ,  $n = 1, \dots, N$ , and running

for  $m = 1, \dots, M$   
do for  $n = 1, \dots, N$

$$\text{do } \Delta_n \leftarrow \arg \min_{\Delta} \{D(S_n + \Delta_n, S_n + \Delta)\}$$

The procedure was used for whole spike trains as well as bursts of the same type, identified by position and structure across multiple repetitions of their associated syllables. To correctly keep the relative timing of spikes within each spike train, in either case the whole spike trains were shifted accordingly. It typically took only a few iterations for 50–80 spike trains to be aligned well enough to reveal their common structures. Then individual common structures were aligned to get finer temporal patterns in the pre-motor activity.

### Analysis of Singing

Singing is a more complex phenomenon than the pre-motor activity of a single RA neuron, and males' songs were occasionally contaminated with cage noises and female vocalizations. Therefore, it was important to select appropriate features to characterize the singing behavior.

The features we used are local peaks in spectrographs (Figures 1A–1C). For high temporal resolution, syllable spectrographs were calculated by FFT's using 128-point Hamming windows, at either 4-point or 1-point steps (20 kHz sampling rate). Let  $A$  be a spectrogram. Given  $\Delta t > 0$  (fixed at 64 samples, or 3.2 ms), at fixed frequency  $f$ , if  $|A(f, t)| < |A(f, t_0)|$  for all  $t \neq t_0$  in  $[t_0 - \Delta t, t_0 + \Delta t]$ , then we define  $t_0$  as the time of a local peak in  $|A(f, t)|$ . (This definition will fail for signals with equal amplitude throughout a small neighborhood within the interval, but in practice, this was never a problem.) To distinguish local peaks in vocalizations from those in background noise, which usually had a low energy level, only peaks with magnitude larger than a threshold were registered. Thresholds were chosen by manual inspection. Different types of syllables were associated with different thresholds, while for syllables of the same type across multiple recording sessions on multiple days, the threshold was fixed. We observed a tendency at higher frequencies for the number of features to decrease and for the scatter in features to increase. Thus, we chose a range of relatively low frequencies for analysis, 1484, 1641, 1797, or  $1953 \pm 78$  Hz for the data presented in Table 1.

Our definition of an acoustic feature implies that most of the features are reliably distributed along the spectral contours of a vocalization, giving compact representations of the global shapes of the spectrographs. Therefore, despite the simplicity of individual features, their aggregate contains information regarding the frequency modulations within syllables, which are the most prominent acoustic features. Since timing of peaks in amplitude is less subject to noise than amplitude itself, we chose analysis of timing as a more reliable method for describing the structure of song.

In accordance with the idea that pre-motor activity regulates singing, first we aligned the neuronal data without any reference to the associated acoustic data, yielding optimal shift  $\Delta_i$  for the  $i^{\text{th}}$  spike train. Then, we shifted each acoustic trace by the  $\Delta_i$  corresponding to the associated neuronal data. We also adjusted for a 40 ms lead in pre-motor neuronal activity. (The exact value of pre-motor lead is not critical; the adjustment merely places the song acoustic feature in the vicinity of the associated neuronal activity.)

### Measuring Feature Alignment

Our analysis of temporal patterns in singing depends on detecting feature alignments and evaluating the acoustic scatter. Suppose  $F_1, \dots, F_N$  are temporal registries of features of  $N$  aligned syllables. With fixed bin width  $\Delta$ , let  $N(t, \Delta)$  be the total number of features in  $F_1, \dots, F_N$  occurring in  $I_t = [t - \Delta/2, t + \Delta/2]$  (Figure 1F). Given fixed threshold  $p$  ( $= 0.7$ ), if  $N(\tau, \Delta) = pN$  and  $N(\tau, \Delta) = N(t, \Delta)$  for all  $t$  in  $I_\tau$ , then  $\tau$  was considered a location where a feature alignment occurs nearby. Let  $S$  consist of the temporal registries of the  $N(\tau, \Delta)$  features in  $I_\tau$ . Then, the location and scatter of the alignment were given by the mean and standard error of the values of  $S$ , i.e.,

$$m = \frac{1}{N(\tau, \Delta)} \sum_s t, \quad (4)$$

$$\sigma = \left( \frac{1}{N(\tau, \Delta) - 1} \sum_s (t - m)^2 \right)^{1/2}$$

The smaller the  $\sigma$ , the tighter the alignment.

For some analyses, we required the temporal registry of the feature occurring in  $I_t$  for each  $F$ . Because of noise in the acoustic data, in some  $F$ 's, features statistically intrinsic to the syllable type were missing, while in some other  $F$ 's, two or more local peaks in amplitude occurred in  $I_t$ , the extraneous peaks resulting from contaminating cage noises or female calling. In the first case, we simply skipped the exemplars, while in the second case, we took the mean location of the local peaks as the location of the "real" feature.

To optimize the localization of the aligned features, we attempted to choose the bin width  $\Delta$  to include at most a single aligned feature. If  $\Delta$  was too large, a bin might include two or more aligned features. We set  $\Delta = 10$  ms, with smaller values for syllables with denser features. Conversely, if  $\Delta$  was too small, this would overestimate the tightness of feature alignment. The following two criteria were set for selecting the lower bound of  $\Delta$ : (1) for at least one  $t$ ,  $N(t, \Delta) = pN$ . This assures a bin width sufficiently large to detect at least one aligned feature. (2)  $\Delta/2 = 2\sigma$  for at least one aligned feature (i.e., the bin covers a 95% confidence interval for the mean temporal location of the aligned feature). Criterion (2) assumes that, conditional on the location of alignment, the times of aligned features are independent and follow the same normal distribution.

To further test whether the bin width criterion was eliminating real variance of acoustic features, we considered a subset of cases characterized by well-separated acoustic features, and minimal contamination of the recordings. In half (6/12) of these cases, doubling the bin width did not alter the number of acoustic features detected, hence the scatter was identical. In five other cases, 1–3 additional features were detected, and in the remaining case, seven additional features were detected. These few additional features all appeared as outliers, apparently caused by noise or female calling, and greatly affected the value of the scatter (95%  $\pm$  27% increase,  $n = 6$ ). This analysis suggests the criterion values we chose did not skew the values obtained for the larger dataset.

Applying these criteria to exemplar syllables yielded multiple acoustic feature alignments within the syllables. For each aligned spike burst, we then found all the acoustic feature alignments within  $\pm 50$  ms from the burst, resulting in a feature count per burst. In

many cases, an acoustic feature occurred within  $\pm 50$  ms of two or more bursts and was counted multiple times. The sum of the counts across all bursts of a unit is reported in Table 1, column 2. For each of these resultant acoustic alignments, we calculated the scatter of the features around that alignment. The mean and the standard error of the scatter across all aligned bursts and the associated acoustic features are reported in Table 1, columns 4 and 5.

### Identifying Songs

The zebra finches in these experiments tended to sing in "bouts," defined as multiple songs produced in close succession followed by longer intervals ( $>30$  s) of silence. Each "song" typically had multiple introductory "notes" followed by one or more "motifs." Each motif consisted of a sequence of "syllables." Syllables typically comprised multiple notes. For each bird, there was at least one highly complex syllable (many notes) in the motif.

By the hierarchical organization of zebra finch singing, identifying songs was reduced to identifying motifs that followed one or more introductory notes. We identified motifs by noting that in our data set, each motif contained a single complex syllable.

### Acknowledgments

We thank Timothy Q. Gentner, Nicholas G. Hatsopoulos, and two anonymous reviewers for helpful comments on the manuscript. Supported by NIH grants MH59831 and MH60276 to D.M.

Received November 15, 2000; revised September 27, 2001.

### References

- Baker, S.N., and Lemon, R.N. (2000). Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels. *J. Neurophysiol.* **84**, 1770–1780.
- Carr, C.E., Heiligenberg, W., and Rose, G.J. (1986). A time-comparison circuit in the electric fish midbrain. I. Behavior and physiology. *J. Neurosci.* **6**, 107–119.
- Date, A., Bienenstock, E., and Geman, S. (1998). On the temporal resolution of neural activity. Technical report, Div. Applied Math, Brown University, Providence, RI.
- Dave, A., and Margoliash, D. (2000). Song replay during sleep and computational rules for sensorimotor vocal learning. *Science* **290**, 812–816.
- deCharms, R.C., and Merzenich, M.M. (1996). Primary cortical representation of sounds by the coordination of action-potential timing. *Nature* **381**, 610–613.
- deCharms, R.C., and Zador, A. (2000). Neural representation and the cortical code. *Annu. Rev. Neurosci.* **23**, 613–647.
- Georgopoulos, A.P., Schwartz, A.B., and Kettner, R.E. (1986). Neuronal population coding of movement direction. *Science* **233**, 1416–1419.
- Hatsopoulos, N.G., Ojakangas, C.L., Paninski, L., and Donoghue, J.P. (1998). Information about movement direction obtained from synchronous activity of motor cortical neurons. *Proc. Natl. Acad. Sci. USA* **95**, 15706–15711.
- Hessler, N.A., and Doupe, A.J. (1999). Social context modulates singing-related neural activity in the songbird forebrain. *Nat. Neurosci.* **2**, 209–211.
- Keat, J., Reinagel, P., Reid, R.C., and Meister, M. (2001). Predicting every spike: a model for the responses of visual neurons. *Neuron* **30**, 803–817.
- Laurent, G., Wehr, M., and Davidowitz, H. (1996). Temporal representations of odors in an olfactory network. *J. Neurosci.* **16**, 3837–3847.
- Lisman, J.E. (1997). Bursts as a unit of neural information: making unreliable synapses reliable. *Trends Neurosci.* **20**, 38–43.
- Mooney, R. (1992). Synaptic basis for developmental plasticity in a birdsong nucleus. *J. Neurosci.* **12**, 2464–2477.
- Oram, M.W., Wiener, M.C., Lestienne, R., and Richmond, B.J. (1999). Stochastic nature of precisely timed spike patterns in visual system neuronal responses. *J. Neurophysiol.* **81**, 3021–3033.

- Pauluis, Q., and Baker, S.N. (2000). An accurate measure of the instantaneous discharge probability, with application to unitary joint-even analysis. *Neural Comp.* *12*, 647–669.
- Searcy, W.A., and Yasukawa, K. (1996). Song and female choice. In *Ecology and Evolution of Acoustic Communication in Birds*, D.E. Kroodsma, and E.H. Miller, eds. (Ithaca, NY: Cornell University Press), pp. 454–473.
- Shadlen, M.N., and Newsome, W.T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* *18*, 3870–3896.
- Singer, W., and Gray, C.M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* *18*, 555–586.
- Sossinka, R., and Böhner, J. (1980). Song types in the zebra finch (*Poephila guttata castanotis*). *Z. Tierpsychol.* *53*, 123–132.
- Spiro, J.E., Dalva, M.B., and 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.
- Suthers, R.A., Goller, F., and Pytte, C. (1999). The neuromuscular control of birdsong. *Phil. Trans. R. Soc. Lond. B.* *354*, 927–939.
- Usrey, W.M., and Reid, R.C. (1999). Synchronous activity in the visual system. *Annu. Rev. Physiol.* *61*, 435–456.
- Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., and Aertsen, A. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature* *373*, 515–518.
- Wild, J.M. (1993). Descending projections of the songbird nucleus robustus archistriatalis. *J. Comp. Neurol.* *338*, 225–241.
- Yu, A.C., and Margoliash, D. (1996). Temporal hierarchical control of singing in birds. *Science* *273*, 1871–1875.