

A Generative Model for Measuring Latent Timing Structure in Motor Sequences

Christopher M. Glaze^{1*}, Todd W. Troyer²

1 Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **2** Department of Biology, University of Texas at San Antonio, San Antonio, Texas, United States of America

Abstract

Motor variability often reflects a mixture of different neural and peripheral sources operating over a range of timescales. We present a statistical model of sequence timing that can be used to measure three distinct components of timing variability: global tempo changes that are spread across the sequence, such as might stem from neuromodulatory sources with widespread influence; fast, uncorrelated timing noise, stemming from noisy components within the neural system; and timing jitter that does not alter the timing of subsequent elements, such as might be caused by variation in the motor periphery or by measurement error. In addition to quantifying the variability contributed by each of these latent factors in the data, the approach assigns maximum likelihood estimates of each factor on a trial-to-trial basis. We applied the model to adult zebra finch song, a temporally complex behavior with rich structure on multiple timescales. We find that individual song vocalizations (syllables) contain roughly equal amounts of variability in each of the three components while overall song length is dominated by global tempo changes. Across our sample of syllables, both global and independent variability scale with average length while timing jitter does not, a pattern consistent with the Wing and Kristofferson (1973) model of sequence timing. We also find significant day-to-day drift in all three timing sources, but a circadian pattern in tempo only. In tests using artificially generated data, the model successfully separates out the different components with small error. The approach provides a general framework for extracting distinct sources of timing variability within action sequences, and can be applied to neural and behavioral data from a wide array of systems.

Citation: Glaze CM, Troyer TW (2012) A Generative Model for Measuring Latent Timing Structure in Motor Sequences. *PLoS ONE* 7(7): e37616. doi:10.1371/journal.pone.0037616

Editor: Melissa J. Coleman, Claremont Colleges, United States of America

Received: January 12, 2012; **Accepted:** April 23, 2012; **Published:** July 16, 2012

Copyright: © 2012 Glaze, Troyer. 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.

Funding: The work was supported by National Institutes of Health grant 5R21MH 066047-02, National Science Foundation grant IOS-0951310 and a fellowship for CMG from the Center for the Comparative Biology and Evolution of Hearing (C-CEBH). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: TWT is a PLoS ONE Editorial Board member. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials.

* E-mail: cglaze@sas.upenn.edu

Introduction

There has been an increasing focus in systems neuroscience on the role of trial-to-trial variability in the neural code for movement [1–3]. Motor variability has been of particular interest in songbird research, including studies on zebra finch song: males exert active control over song variability, singing more stereotyped song when courting a female than when singing alone [4–7]. Further experiments have shown that this control is largely mediated by the anterior forebrain pathway, a cortical-basal ganglia circuit for song [5–8].

Interestingly, single unit recordings in singing birds reveal that bursting in individual neurons is locked to song output with millisecond precision [9–11], with trial-to-trial variability in interburst intervals being strongly correlated with variability in the timing of corresponding song features [10]. Several studies have exploited this relationship and used precise measurements of song timing to make inferences about the nature and timescale of the underlying motor code [12–15].

However, studies aimed at linking brain and behavior face the challenge that variability is likely to be driven by multiple sources with different timescales and mechanistic explanations. For example, a natural way to measure timing variability would be to measure the standard deviation of individual song elements known as “syllables.”

While some of that timing deviation is likely due to fast neural and synaptic noise within the central circuits that generate song timing, variability in syllable length is also likely to be affected by slower modulations that alter song tempo. Finally, alterations in timing introduced downstream from the song pattern generator, either in motor periphery or by measurement errors in syllable duration, will contribute to variability in measured syllable duration as well. To understand the mechanisms of song production one would like to separate these distinct components of timing variability.

We have developed a statistical model to address this problem, allowing researchers to distinguish three kinds of timing variability in repeated action sequences such as zebra finch song: global timing changes spread across the sequence (“tempo”), fast, uncorrelated timing noise in the length of each element of the sequence (“independent variability”), and timing “jitter” at the onsets and offsets of sequence elements that might be caused by variation in the motor periphery or by measurement error. This “timing variability model” can be seen as an extension of the Wing-Kristofferson model [16], which separates central and peripheral components of variability in human interval timing during a finger-tapping task. Our approach uses a novel variation on factor analysis, a technique for detecting and measuring latent variables in high-dimensional data sets [17–19]. We fit the model

to data using an expectation-maximization (EM) algorithm, which allowed us to estimate the contribution of each of the three latent factors to the variability of each element of the sequence.

We applied the model to the songs of 11 adult male zebra finches and found that for most birds the model provides very good fits to trial-to-trial variability in the durations of individual song vocalizations (syllables) and the gaps of silence between them. For individual syllables and gaps, a roughly equal amount of variability is contributed by each of the three components of timing variability. For 5 birds, the model appears to be optimal when including an additional latent variable in which syllable and gap durations are anti-correlated. We also performed Monte Carlo simulations to investigate how accurately the model was able to fit artificial data sets where the underlying timing parameters and latent variables were known.

Results

This paper develops a statistical model to estimate latent factors governing trial-to-trial timing variation in repeated action sequences. We will present our results in three sections. First, we describe our model which we call the “timing variability model.” We then fit the model to timing data from zebra finch songs, and present several results regarding the nature of timing variability in this species. Finally, we describe Monte Carlo simulations designed to quantify the reliability of the model fits.

Timing Variability Model

The timing variability model can be applied to data consisting of repeated action sequences that can be divided into consecutive time intervals, as might be observed during the production of a stereotyped behavior over multiple trials. Although the model may be applied to any action sequence, here we apply the model to quantify timing variability in the songs of male zebra finches. Each adult produces a song that is highly invariant but unique to that male (Fig. 1A). The song of each male consists of a stereotyped string of 3–7 distinct vocalizations we term “syllables” that vary in duration from approximately 50 to 200 msec; syllables are separated by gaps of silence that are roughly half the duration of syllables. We will use the term “motif” to denote a male’s stereotyped string of syllables, while a song “bout” consists of a series of motifs produced back-to-back. We restricted the analysis to the first motif of each bout; however, to examine the gap of silence between the 1st and 2nd motifs, we included the first syllable of the second motif in the sequences analyzed. We will use the term “time interval” to denote the durations of both syllables and silent gaps, so if a sequence consists of L syllables, it also contains $L - 1$ gaps and thus $2L - 1$ time intervals. The duration of each syllable and silent gap is also fairly stereotyped, varying across motif renditions by about 1.5–3.5 msec standard deviation in duration, or ~ 2 –5% coefficient of variation (CV) [15]. This temporal stereotypy gives each male’s song a very characteristic rhythm [15,20].

Variance decomposition. To start we describe the model from the perspective of variance decomposition, generally following the standard approach known as factor analysis [17,19,21,22]. As outlined in the introduction, we would like to write the variance in interval length as a sum of three component variances: variance in tempo spread across the song, independent variance that is specific to each timing interval within the song, and timing jitter in the exact placement of the onset and offset of each song syllable.

For the first factor, we presuppose a non-observable, latent factor \mathbf{z} that influences song tempo, with \mathbf{z}_n representing the

deviation in tempo from its mean value on each trial n . If the change in tempo is spread uniformly over the song, the deviation in the length of interval k will be proportional to the mean length of that interval, \bar{x}_k ; that is, the deviation will equal $\mathbf{z}_n \bar{x}_k$. However, the timing variability model is not restricted to uniform changes in tempo. Rather, deviations in global timing are modeled as $\mathbf{W}_k \mathbf{z}_n$ where the values \mathbf{W}_k are free parameters representing the relative sensitivity of each interval to variations in global timing. For that reason, we refer to \mathbf{W}_k as the “global weight” for interval k . Under these assumptions, variations in global timing will contribute a variance component to interval k equal to $\mathbf{W}_k^2 \text{Var}(\mathbf{z})$ where $\text{Var}(\mathbf{z})$ represents the variance of the latent factor \mathbf{z} . Importantly, global timing will also result in a covariance between interval j and interval k equal to $\mathbf{W}_k \mathbf{W}_j \text{Var}(\mathbf{z})$. By fitting this pattern of covariances to the off-diagonal elements of the data covariance matrix, the model forms an estimate of the global sensitivity of each interval to global timing (Fig. 1C, left).

A similar approach can be used to estimate the variance contributed by timing jitter. For this component, we presume that the length of each interval is set centrally, but the measured timing of the transition between consecutive intervals is subject to variation downstream of central pattern generator for song, either in the motor periphery or during the measurement process. We let \mathbf{u}_{kn} denote the jitter in timing between interval k and $k + 1$ on trial n . As noted by Wing and Kristofferson [16] in the context of human finger-tapping experiments, such timing jitter introduces a negative covariance between adjacent intervals in addition to contributing to the variance in interval duration (Fig. 1C, right). As before, off-diagonal values in the data covariance matrix can be used to estimate the magnitude of timing variance at each boundary in the sequence. We denote the variance at the k th boundary by Ω_{kk} . The variance contributed by jitter the timing of interval k will be $\Omega_{kk} + \Omega_{k-1,k-1}$.

Finally, we assume that each interval k is subject to its own degree of independent variation η_{kn} on trial n . This independent variance can be estimated by calculating the total interval variance and subtracting the variance contributed by global and jitter factors. We denote the independent variance for interval k by Ψ_{kk} .

Variance decomposition in the model is shown in figure 1. Panel B shows the data covariance matrix from a single bird. Panel C shows the three matrices of covariance contributed by the global, independent and jitter terms in the model to the right and how these sum to equal the model covariance matrix at the left.

The difference between the timing variability model and traditional factor analysis stems from the explicit incorporation of the negative covariances between adjacent intervals caused by timing jitter at syllable onsets and offsets. It should be noted that since the model can only estimate jitter by recognizing negative covariation between adjacent intervals, estimates of timing variability in the first and last intervals of the sequence will be inherently less accurate. As a result, we discarded the first and last intervals in the sequence from all reported analyses.

Generative model and maximum likelihood estimation. We used a maximum likelihood approach to fitting our model to the data [17,22,23]. More specifically, we view the timing variability model as a probabilistic model capable of generating artificial data consisting of N repeated sequences, each containing the duration of K consecutive time intervals. To generate such data, we assume that the parameters of the model are fixed and for each artificial sequence n we generate random values for the unobserved, latent variables: a single global timing variable, \mathbf{z}_n , independent variations for each of K intervals, η_{kn} , and jitter variations at each of the $K - 1$ boundaries between intervals, \mathbf{u}_{kn} . These are combined to produce K observed interval durations

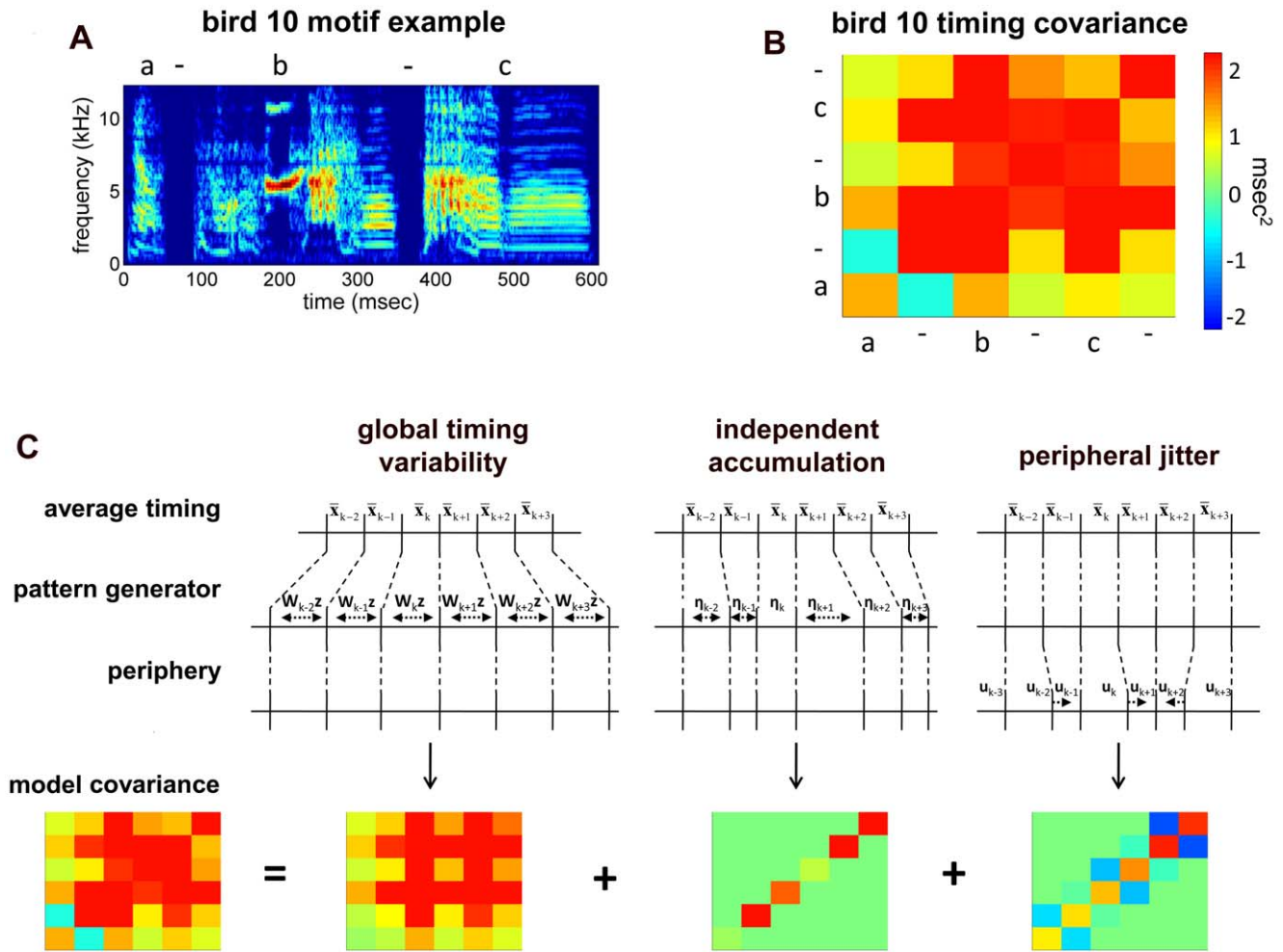


Figure 1. Timing variability and its decomposition. A, Spectrogram of a single song motif from one bird, with syllables indicated above by letters and silent gaps by “-”. B, timing covariance matrix for the same bird, in which the color of each square indicates the pairwise covariance between the durations of two intervals denoted along the rows and columns. C, Schematic of the different noise sources captured in the model. Below each noise source is a covariance matrix generated by the model of that source. doi:10.1371/journal.pone.0037616.g001

$$\mathbf{x}_{kn} = \mathbf{W}_k \mathbf{z}_n + \eta_{kn} + \begin{cases} -\mathbf{u}_{k-1,n} + \mathbf{u}_{kn} & 1 < k < K \\ \mathbf{u}_{1n} & k = 1 \\ -\mathbf{u}_{K-1,n} & k = K \end{cases} \quad (1)$$

For convenience we can rewrite this equation in vector notation as:

$$\mathbf{x}_n = \mathbf{W} \mathbf{z}_n + \eta_n + \mathbf{D} \mathbf{u}_n \quad (2)$$

where \mathbf{D} is the $K \times (K - 1)$ differencing matrix with ones along the diagonal ($\mathbf{D}_{k,k} = 1$ since boundary k causes deviations in the offset of interval k), negative ones along the subdiagonal ($\mathbf{D}_{k,k-1} = -1$ since boundary $k - 1$ causes deviations in the onset of interval k), and 0 elsewhere.

To allow theoretical derivations of maximum likelihood fits, we assume that the latent factors are independent and Gaussian distributed. The free parameters of the model are then the global weights in \mathbf{W} (collectively referred to as the global parameters) and the variances of η_k and \mathbf{u}_k . We let Ψ and Ω be the covariance matrices of the random vectors η and \mathbf{u} respectively; these matrices

are diagonal since the latent variables in η and \mathbf{u} are assumed to be independent. The variance of the global parameter \mathbf{z} is fixed at a value of one, with the magnitude of the global weights \mathbf{W} determining the magnitude of the observed variances. Letting \mathbf{S} be the covariance matrix of interval durations, and using the basic result that $\mathbf{S} = \frac{1}{N} \sum_n \mathbf{x}_n \mathbf{x}_n^T$, equation 2 implies that in the limit of large N , the timing variability model will generate K dimensional Gaussian distributed data with covariance matrix

$$\mathbf{S} = \mathbf{W} \mathbf{W}^T + \Psi + \mathbf{D} \Omega \mathbf{D}^T \quad (3)$$

In particular, for any given sequence of interval durations, we can calculate the probability that the timing variability model would generate that particular sequence.

To fit the model to observed data, we find parameters \mathbf{W} , Ψ and Ω that maximize the likelihood that the model would generate the observed collection of sequence durations. To accomplish this, we use an Expectation-Maximization (EM) algorithm that was modified from the standard EM algorithm used to generate maximum likelihood fits for standard factor analysis [17,23]. (See Methods for details.) A consequence of this maximum likelihood,

generative approach is that the EM algorithm naturally yields estimates of the latent variables \mathbf{z}_n , \mathbf{u}_n and η_n on a trial-to-trial basis. As we will demonstrate below, this allows for an analysis of the relationship between timing factors and other variables, such as time of day and behavioral state, that vary across samples but are not explicitly included in the model. Furthermore, the model can be used to generate artificial data sets from a model with known parameters. These artificial data sets can then be fit using the same class of model. This exercise allows one to estimate the uncertainty in parameter estimates that are due solely to statistical variation in finite data sets (see ‘Monte Carlo experiments’ below).

Timing Variability in Zebra Finch Song

We applied the timing variability model to song samples from 11 adult zebra finch males reported in a previous study [15]; one bird from the previous study was omitted because the sample was not large enough to yield reliable fits. For each male we collected a sample of 215–885 song bouts recorded while the bird was alone, housed with another adult male, or serving as a song tutor to a juvenile. Data from individual birds often included recordings from multiple recording sessions, spanning periods that ranged from 1–6 months. We restricted the analysis to the first motif of each bout; however, to examine the gap of silence between the 1st and 2nd motifs, we included the first syllable of the second motif in the sequences analyzed. We will refer to gaps between motifs as “inter-motif gaps” and distinguish them from “within-motif gaps”. After fitting the model to the data, we then excluded the first and last syllable of each sequence since the parameter estimates for these intervals are corrupted by unknown jitter at the beginning and end of the sequence (see above). The resulting data set included parameter estimates for 48 syllables, 48 within-motif gaps and 11 between-motif gaps, for a total of 107 time intervals. Throughout we will report parameter distributions with median \pm median absolute deviation (MAD).

Model fit. We assessed the overall fit of the model to the data using the standardized root mean-squared residual (SRMR) [24]. This measure is based on the average difference between the pairwise interval length correlations in the data and those predicted by the model covariance matrix (see Methods for formula). As a rule of thumb in factor analysis, a SRMR < 0.08 is considered an adequate fit while SRMR < 0.05 is considered an excellent fit [24]. Across birds, SRMR was a median 0.058, range of 0.034–0.105. Nine of eleven birds yielded an SRMR < 0.08 while 3 yielded SRMR < 0.05 (Fig. 2A). Thus, the model fit a large majority of birdsongs in our sample reasonably well. The particularly large SRMR values appeared to be due to timing factors not included in the original model but analyzed below.

Parameter distributions and the contribution of tempo. Across all time intervals, median global variability was 0.972 ± 0.391 msec, median independent variability was 1.110 ± 0.527 msec, and median boundary jitter was 0.921 ± 0.354 msec. Averaging across intervals produced by a given bird, respective medians ranged from 0.550–1.456 msec (global), 0.560–2.348 msec (independent), and 0.638–1.723 msec (jitter). Thus, each of the three timing factors made similar contributions to the variation in duration of individual song elements. In general, timing variability was relatively small in these songs. It may also be noted that, when respective variances are summed to estimate total duration variability in song elements (assuming an interval receives jitter variance from two sources), the numbers agree with previously reported data: ~ 1 msec² variance in each component yields ~ 4 msec² interval variance total (so ~ 2 msec st. dev.), on par with what we have previously reported [15].

Based on a previous analysis showing generally positive correlations among the durations of time intervals, we expected global timing variations in which most time intervals shortened or lengthened together [15]. Of the 107 song intervals analyzed, only 4 had global weights < 0 , with two of the negative weights coming from a single bird (out of 11 time intervals for that bird). Since nearly all global weights were positive, global timing variations indeed were dominated by changes in which time intervals shortened or lengthened together. Therefore, the global factor represents changes in a shared tempo across the song and we will use the terms “global variability” and “tempo variability” interchangeably.

Although global weights were on the same order of magnitude as the independent variability for individual intervals, the variance in total motif length was dominated by tempo changes. Because global tempo is shared by all song elements, the motif length variance linked with the global variable (equal to $(\sum_k \mathbf{W}_k)^2$) was roughly ~ 5 times as much as the length variance linked with the independent variable (equal to $\sum_k \Psi_{kk}$). More specifically, tempo explained $83.2 \pm 6.5\%$ of the total variance in motif duration explained by the model (range 31.0–91.2% by bird). Only one bird (bird 4) had more motif length variance linked with the independent variable than with the global; interestingly, this appeared to be entirely due to independent variance within a silent gap corresponding to a variable sequence point at which the bird occasionally transitioned to a different syllable.

Overall, the motif length variance in the model (equal to the sum of global and independent variance) deviated from measured motif length variance by only $0.245 \pm 0.240\%$ across birds (maximum of 3.6%).

Length scaling. A wide range of studies, from human finger-tapping experiments to previous work in birdsong, have found that variability in the duration of a learned motor gesture tends to scale with the average duration of the gesture [13,15,16,25–28]. This is to be expected from any process that consists of a causal sequence of activity with noise at each time point: the longer the process, the more variability accumulates. We hypothesized that variability would scale with average duration for the tempo and independent timing parameters, as these parameters are expected to alter timing throughout the duration of a given interval. In contrast, we did not expect any relationship between average duration and the variability linked with boundary jitter. To quantify the boundary jitter linked with interval k , we calculate the square root of the sum of jitter variances at the two ends: $\sqrt{\Omega_{kk} + \Omega_{k-1,k-1}}$.

Indeed, the data confirmed that only tempo-based and independent variability scale with average interval duration (Fig. 3): the correlation between global variability and average interval duration was 0.701 (Spearman, $p < 0.0001$), between independent variability and duration, 0.293 (Spearman, $p < 0.005$). In contrast, average duration showed no significant correlation with jitter-based variability (Spearman’s $r = 0.060$, $p = 0.54$). Although the relationship between independent variability and average duration was significant, it was far weaker than what we found with tempo-based variability. There are a number of reasons why this might be, including greater error in independent parameter estimates as well as real factors in the neural system that may contribute noise in a way that is not dependent on process duration.

Model parameters distinguish syllables from silent gaps. Previous research has shown that silent gaps stretch and compress with song tempo more than syllables after dividing out average duration, *i.e.* they are more “elastic” to trial-to-trial tempo variability [13,15] (although see [12]). We examined whether the global parameters from our model indicated the same difference

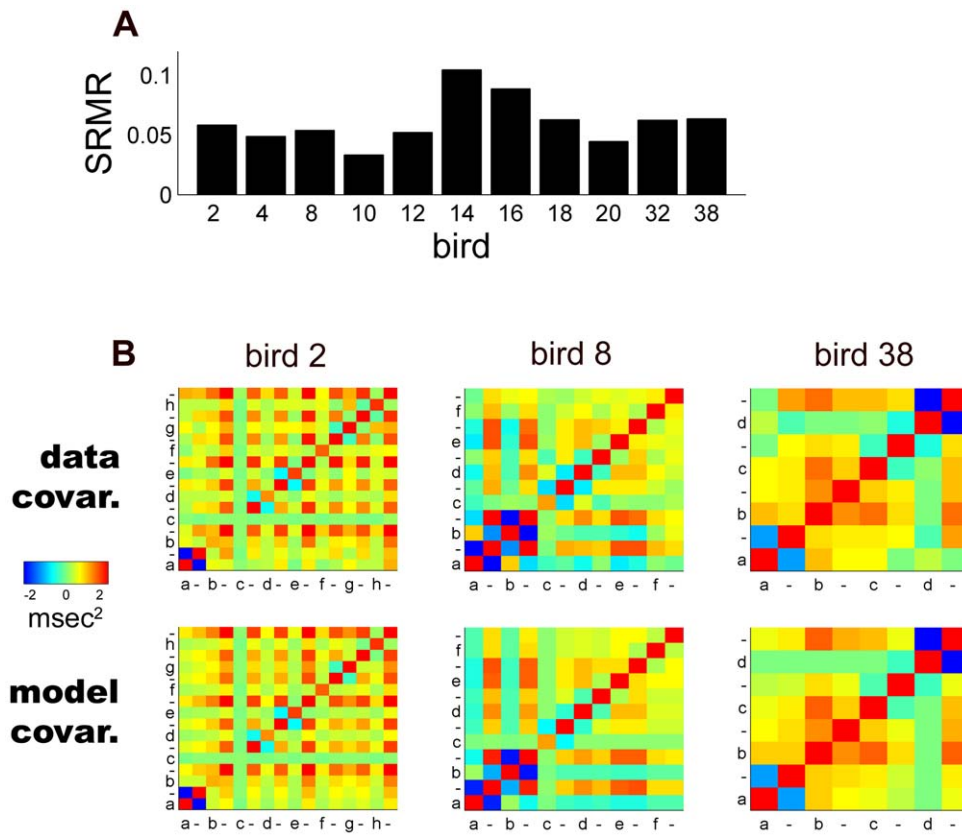


Figure 2. Goodness of fit between the timing model and data from 11 zebra finch birds. A, SRMR by bird (see Results). B, Color plots of the timing covariance matrices from the data and model for 3 representative birds. doi:10.1371/journal.pone.0037616.g002

between syllables and gaps. Here, we computed tempo elasticity for interval k as $\mathbf{W}_k/\bar{\mathbf{x}}_k$, where $\bar{\mathbf{x}}_k$ is the mean duration of the k th time interval. This measure may be viewed as a partial CV with respect to tempo so we will occasionally refer to it as “global CV”. For example, a global CV of 0.01 in an interval would indicate

that the standard deviation of global timing expressed in the interval is 1% of its average duration.

Indeed, the model verified that syllables were less elastic to tempo than gaps (Fig. 3A), with most gaps falling above the regression line of global variability on mean duration and most syllables falling below. Syllable tempo elasticity was 0.011 ± 0.003 ,

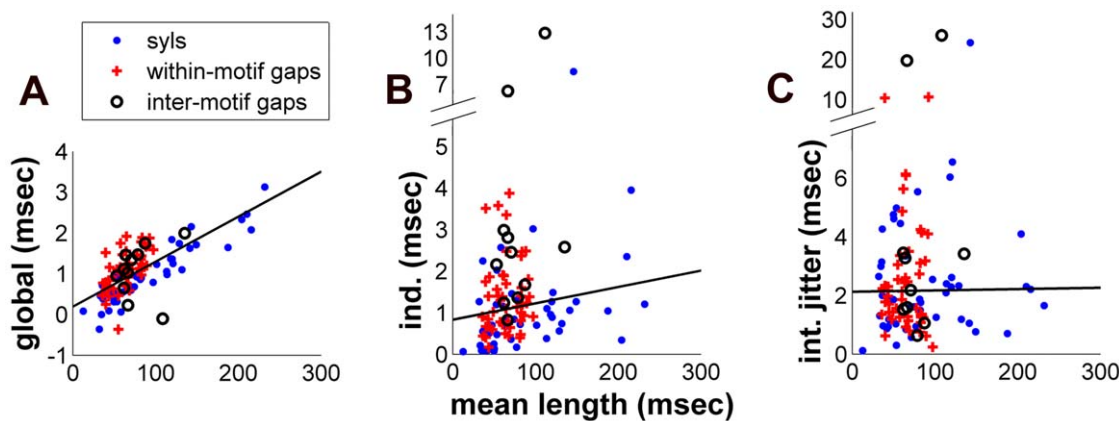


Figure 3. Scaling of timing parameters with average length. A, Scatter plot of global tempo variability by average interval duration across all syllables and silent gaps. Silent gaps within motifs and between motifs have been separated out. B, C, Analogous plots for the independent and jitter parameters respectively. In each plot solid lines come from a regression of timing parameters on average interval duration, calculated across birds. Overall, plots show that global and independent variability scale with average duration, while jitter does not. The plots also show systematic differences among interval types with respect to how much each parameter scales with average duration. doi:10.1371/journal.pone.0037616.g003

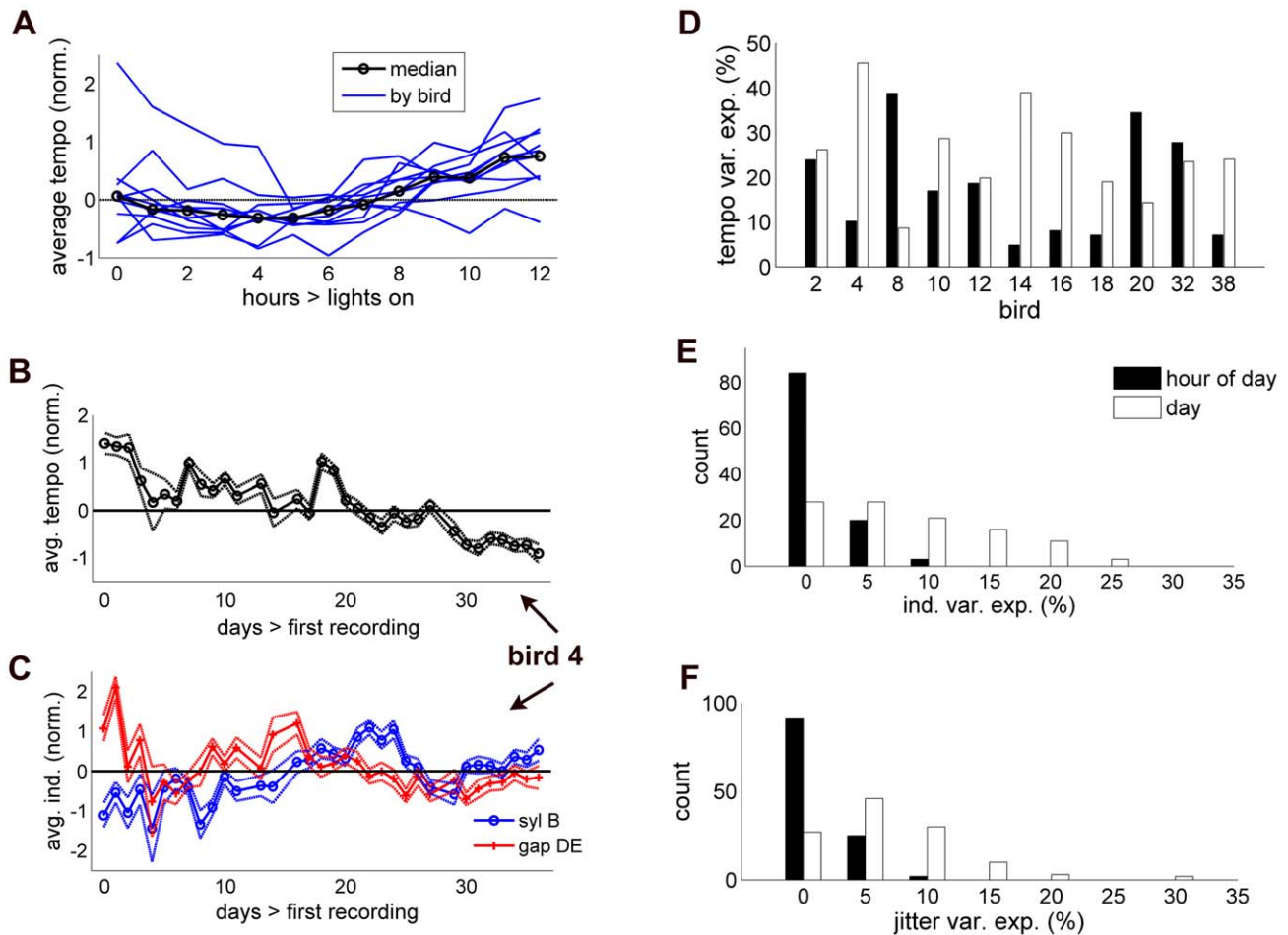


Figure 4. Circadian rhythm and day-to-day drift in the latent timing variables. A, Average tempo by hour of the day (defined from lights on) by bird (blue lines) and averaged across birds (black). B, example of day-to-day drift in tempo for one bird over a 1-month period. C, example of drift in independent variability for two different song segments in the same bird. All data in plots A-C are group means from a two-factor ANOVA that have been adjusted for unequal samples of factor combinations (see Results), while dotted lines in B and C indicate adjusted means \pm standard error. D, The estimated amount of variance in tempo explained by both hour of the day (black) and day (white) across all 11 birds. E and F, histograms of the amount of variance in the independent and jitter variables respectively, explained by both hour of the day and day (same color code). Generally the data show significant circadian variation in tempo only, and day-to-day drift in all three latent timing variables. doi:10.1371/journal.pone.0037616.g004

while gap tempo elasticity was about 1.5 times greater on average, 0.017 ± 0.003 ($p < 0.0001$, Wilcoxon rank-sum test).

Although previous research has focused on elasticity with respect to shared, tempo-based variability, our timing variability model also allowed us to investigate the pattern of sensitivity to independent variability as well. (We did not analyze jitter differences because syllables and gaps share the same set of interval boundaries and there is no relationship with average duration to begin with.) Because the independent parameters scaled with length we derived an analogous quantity that may be viewed as a partial CV with respect to independent variability, $\sqrt{\Psi_{kk}}/\bar{x}_k$. Interestingly, here we found that as with tempo CV, the independent CV was smaller for syllables than for gaps (Fig. 3B): independent CV was 0.009 ± 0.005 among syllables, and about twice as great on average among gaps, 0.019 ± 0.008 ($p < 0.0001$, Wilcoxon rank-sum test).

Inter-motif gaps express more independent variability. In a hierarchically arranged motor code the onsets of entire motifs may be triggered by mechanisms that are different from what triggers the onsets of individual syllables. There is electrophysiological evidence for this segregation [29], while

previous analysis has suggested increased timing variability at motif boundaries [15]. Therefore, we examined whether the variability in duration of silent gaps between motifs differed from the gaps between syllables within a motif (Fig. 3A–B). Indeed, the 11 inter-motif gaps we investigated had a significantly greater independent CV, 0.035 ± 0.016 , vs. 0.017 ± 0.007 among within-motif gaps ($p < 0.05$, Wilcoxon rank-sum test). Interestingly, we did not find the same difference with respect to global CV: inter-motif gaps had a median 0.018 ± 0.006 vs. 0.017 ± 0.005 among within-motif gaps ($p = 0.54$, Wilcoxon rank-sum test).

Thus, the model revealed timing properties that systematically vary by interval type, with gaps expressing proportionally greater tempo changes and independent variability, and inter-motif gaps expressing an especially large amount of independent variability.

Circadian Influence and day-to-day Drift in Latent Timing Variables

The analyses above examined the relative magnitudes of three components of timing variability and their distribution across song elements. One advantage of using the EM algorithm is that it

naturally provides maximum likelihood estimates of the latent variables on a trial-by-trial basis. This allows investigations in which these unobservable, latent factors are correlated with other biologically relevant variables. For example, changes in song timing can be tracked over periods ranging from hours to days and even months.

Here, we investigated whether any of the latent timing variables underwent consistent patterns of modulation over the course of the day [15] or drifted over the 1–6 months of song recording. Specifically, for each latent variable and each bird we used a Bonferroni corrected two-factor ANOVA to test for significantly different averages by both hour of the day and day of song production. If we found significant differences, we then examined the pattern of changes related to a given factor by calculating a mean and standard error of the latent variable; we adjusted these estimates for unequal sample numbers across factor combinations using the “multcompare” function Matlab.

All 11 birds showed significant tempo changes that were affected by both time of day and day of singing (Fig. 4A,D), indicating both a circadian variation and slow drift in song tempo (Bonferroni-corrected $p < 0.05$). Across birds, the amount of variance explained by hour of the day was a median $17.0 \pm 9.9\%$ (range 4.9–38.8% by bird), by daily drift, $24.1 \pm 5.1\%$ (8.7–45.6%). Consistent with our previous study [15], the pattern of circadian variation in tempo was similar across birds (Fig. 4A), with songs speeding up until mid-morning and slowing down over the afternoon.

The model also allowed us to test for circadian changes and day-to-day drift in the independent and jitter latent variables η and \mathbf{u} (Fig. 4B,C,E,F). In contrast to tempo, neither showed consistent changes by hour of the day, with only 5 of 107 independent components and 3 of 118 jitter components showing significant change by that factor (Bonferroni-corrected $p < 0.05$). Interestingly, many intervals did show significant drift in those variables, with 54/107 independent and 50/118 jitter components reaching significance (Bonferroni-corrected $p < 0.05$). Of the intervals with significant drift, that factor explained $13.0 \pm 4.1\%$ of the independent variation (range 3.1–23.4%) and $9.7 \pm 2.4\%$ of the boundary jitter (range 3.5–30.5%).

Thus, the data indicated day-to-day drift in all three types of latent timing variables, and a significant circadian pattern in tempo only.

Multiple Global Timing Factors

Our original construction of the model incorporated a single global factor, motivated by an a priori expectation of variation in song tempo which was confirmed by the data. However, the factor analysis framework also allowed us to search for additional global factors whose influence is spread across the song sequence. The existence of such additional factors was suggested by a qualitative examination of the data from several birds as well as correspondingly high SRMR values for those model fits. The EM algorithm easily accommodated the inclusion of multiple global factors simply by increasing the number of dimensions in \mathbf{z} and \mathbf{W} and fitting parameters as before (see Methods).

By adding parameters to our model, we are assured of improving the fit to the data. To avoid over fitting, we ran models with 1–4 global dimensions, and used the minimum Bayesian Information Criterion (BIC; see Methods) to determine the optimal tradeoff between better fits to the data and model complexity. Across our sample, 6 of 11 birds had a BIC lowest for just 1 global factor. Of the remaining 5 birds, 4 produced models that had a lowest BIC for 2 global factors while the remaining bird had 3 global factors. As expected, additional global factors

improved the fit to the data: of the 5 birds with > 1 global factor, median SRMR dropped from 0.064 to 0.021 (Fig. 5A); across all birds, median SRMR after using the BIC was 0.038, range 0.016–0.063.

As in traditional multi-dimensional factor analysis, the model fit can only determine the subspace spanned by the matrix \mathbf{W} . Within this subspace there are an infinite number of variations of \mathbf{W} that will give the same pattern of covariances in $\mathbf{W}\mathbf{W}^T$. Based on our previous analysis with one global factor, we sought a solution that separated out tempo-based variability from any additional timing factors. In particular, we used a transformation of \mathbf{W} in which global factors affecting total sequence length would be confined to the 1st dimension, with additional timing factors chosen orthogonal to this tempo dimension (see Methods). In the one bird with 3 global factors, we chose as the 2nd factor the one which explained the most interval variance after factoring out the tempo dimension. We then analyzed the distributions of parameters from the 2nd global factor across the 5 birds with > 1 global dimension.

The second factor tended to have weights of the opposite sign across syllables and silent gaps (Fig. 5B,C). Since factors are ambiguous with respect to overall sign, the second global factor was chosen so that the sum of the syllable weights was positive. With that convention, 26 of 30 syllable weights were > 0 and 23 of 25 gap weights were < 0 . Median syllable weight was 0.329 ± 0.219 msec, vs. -0.389 ± 0.274 msec across gaps ($p < 0.0001$, Wilcoxon rank-sum test), and the difference was consistent across all 5 birds. Interestingly, this pattern corroborates a previous finding of significant length correlations in syllable-syllable and gap-gap pairs after subtracting out the influence of tempo [15]. Unlike the tempo-based weights, we found no relationship between magnitude of the weight for the second component and the mean length of the interval (Spearman's $r = -0.14$, $p = 0.30$). Thus the second global factor showed properties that were similar to jitter: it yielded a tradeoff between syllable and gap lengths and did not scale with average interval duration.

In all 5 birds, the second global variable showed a significant daily drift, and there was a significant circadian component in 3 of 5 birds (Bonferroni-corrected two-way ANOVA, $p < 0.05$; Fig. 5D,E). Hour of day explained 25.0–50.4% (median 28.5%) of the variance in the 3 birds for which the circadian component was significant, while daily drift explained 39.1–69.0% (median 45.1%) across the 5. The trend in the circadian component was for syllables to get longer and the expense of gap duration over the afternoon. Overall, these longer timescales explained a remarkably large amount of variance in the component.

Finally, we examined whether the inclusion of additional global factors significantly changed any parameter estimates from the original model with only 1 global dimension. Among the 5 birds with > 1 global factor included from the BIC analysis, tempo-based weights changed only slightly, with the median decrease of 0.003 msec failing to reach significance ($p = 0.136$, Wilcoxon signed-rank test). The median absolute magnitude of the change was 0.015 ± 0.008 msec, much smaller than the median weight of 0.972 msec. Independent parameters changed more, with median absolute change of 0.049 ± 0.040 msec. However, as with the tempo weights, the direction of change was variable so that the median change of -0.003 msec over all intervals was not significant ($p = 0.64$, Wilcoxon signed-rank test). Jitter parameters also showed a larger absolute change of 0.042 ± 0.038 msec. But unlike independent parameters, jitter parameters tended to get smaller, with 34 of 50 boundaries showing decreased jitter, resulting in a small but statistically significant median decrease of

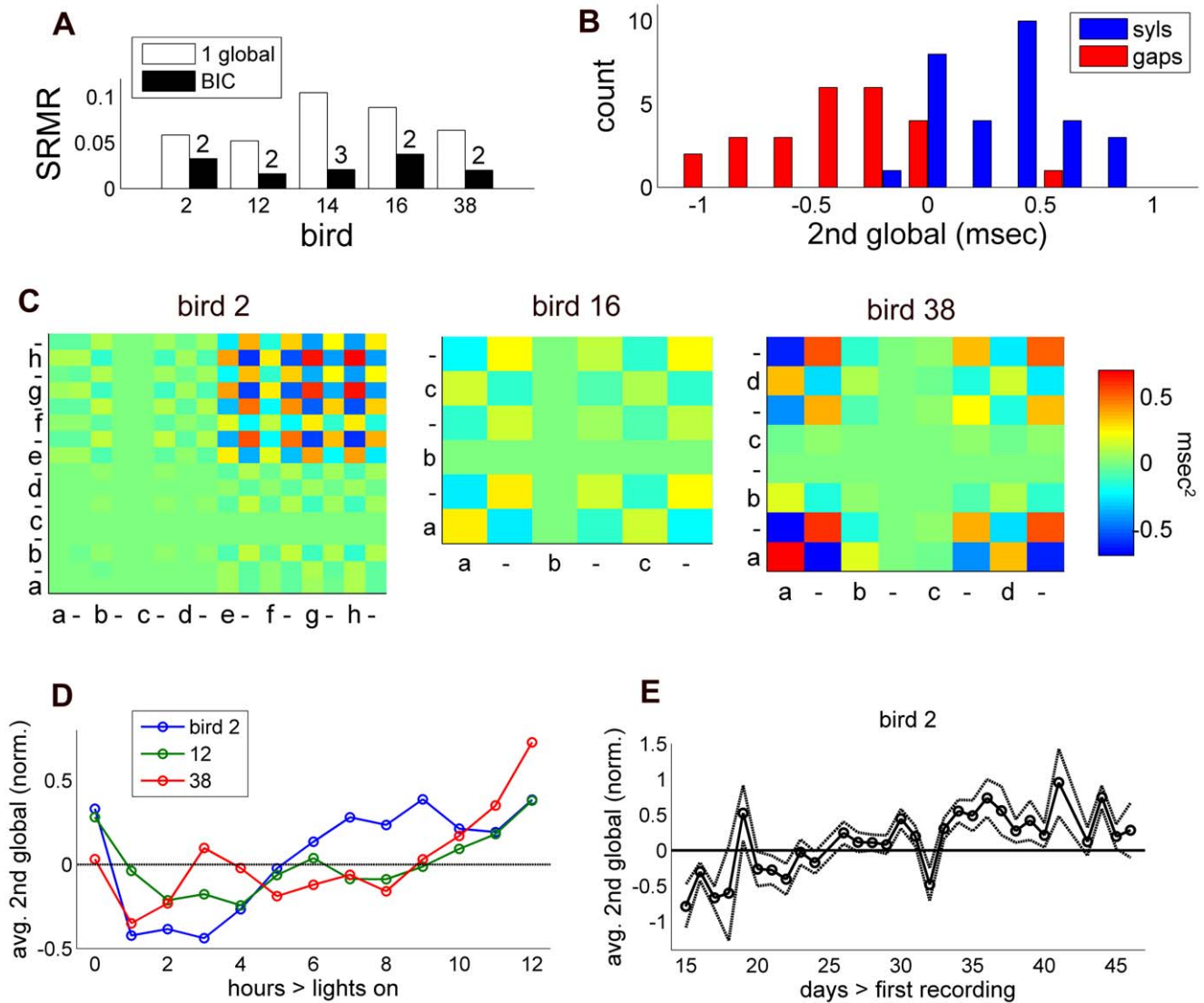


Figure 5. The influence of additional global factors on model fit and song timing. A, Goodness of model fit with additional global factors included (black) vs. 1 global factor (white) for the 5 birds for which BIC was lowest for > 1 factor. Numbers above black bars indicate the number of factors associated with the lowest BIC for that bird. B, Distribution of global parameters for the 2nd factor (sign-normalized, see Methods), separated by whether the song segment was a syllable or silent gap. Across birds syllable and gap parameters tended to be of opposite sign. C, Timing covariance matrices generated for the 2nd global factor in 3 representative birds. D, Average changes in the 2nd global variable in the 3 birds that showed significant circadian variation for that factor. The circadian pattern showed that syllables tended to elongate at the expense of gaps over the afternoon. E, Example of day-to-day drift in the 2nd global variable for one bird over a 1-month period. As in figure 4, hour-of-day and daily averages in D and E have been adjusted for unequal sampling of factor combinations while dotted lines in E represent adjusted means \pm standard error. doi:10.1371/journal.pone.0037616.g005

-0.027 msec ($p < 0.0005$, Wilcoxon signed-rank test). This suggests that a small amount of the estimated boundary jitter in the single-global factor model may be explained by the song-wide anti-correlation between syllables in gaps that was uncovered in the 2nd global factor.

Monte Carlo Experiments

One of the difficulties in using latent variable models is that the model fit cannot be directly validated by comparison with observed data. As an alternative, we investigated how accurately the model was able to fit artificial data sets where the underlying timing parameters and latent variables were known. Specifically, for each bird we selected parameter values that were fit to the experimental data. Then using the model in generative mode, we

created an artificial data set with the same number of samples by randomly generating latent variables from a unit normal distribution, scaling the independent and jitter variables by the corresponding parameters, and combining these into observed interval lengths according to equation (1). Focusing exclusively on this artificial data set, we reapplied our algorithm to find the parameters that provided the best fit to the artificially generated data. In order to derive statistics of how well the model fit the artificial data, we generated 200 distinct artificial data sets for each bird.

These Monte Carlo simulations allowed us to evaluate the reliability of the model in two distinct ways. First, we compared the overall fit of the timing variability model to the real vs. artificially generated data by comparing the model covariance calculated

from equation (3) to the data covariance matrix for both the artificial and real data sets. In all birds, SRMR was smaller for the fit to the simulated data than to the actual data (Fig. 6E). Median SRMR computed across simulations ranged from 0.009–0.054 by bird, roughly two thirds of what we found in the real data. This suggests that about two thirds of the difference between real and model covariance matrices resulted from idiosyncratic components of the covariance matrix due to random sampling, while a third of the difference resulted from patterns in the data that did not conform to the assumptions of the model. It is unclear whether these patterns represent additional global timing factors, specific dependencies between the factors that are not included in the model, or limitations due to other model assumptions.

Second, we compared the parameters and latent variables from the best-fit model to the (known) parameters and latent variables in the artificial data set. This allowed us to assess the reliability of model estimates due purely to statistical variation in the timing of interval resulting from the posited latent factors. We report reliability using the median absolute deviation of estimates from the true values. In the 5 birds with multiple global parameters, we

generated artificial data for both the single-global and multiple-global factor models; however, because error was similar across both approaches (see below) we focus on the multiple global factor models in order to include estimates of the 2nd global factor.

All parameters had similar magnitudes of error (Fig. 6F–I): across bird simulations, tempo parameter error was 0.072 ± 0.026 msec, while error in the 2nd global parameter was a 0.056 ± 0.012 msec; independent parameter error, 0.093 ± 0.028 msec; and jitter error, 0.062 ± 0.025 msec. In the five birds with multiple global factors, the average error was similar in the single-global factor model, with no significant differences among independent and jitter parameters ($p=0.10$ and 0.44 respectively, Wilcoxon signed-rank test), and a small albeit significant decrease of 0.002 msec in error in the tempo-based parameters ($p<0.0005$, Wilcoxon signed-rank test). We also verified that the error in parameter estimates reliably decreased with the square root of sample size: For 3 representative birds we ran 200 simulations at each of several sample sizes ranging from 50 to 5000 ($N=50,100,200,500,1000,2000,5000$). A regression of median parameter error with the square-root of sample size

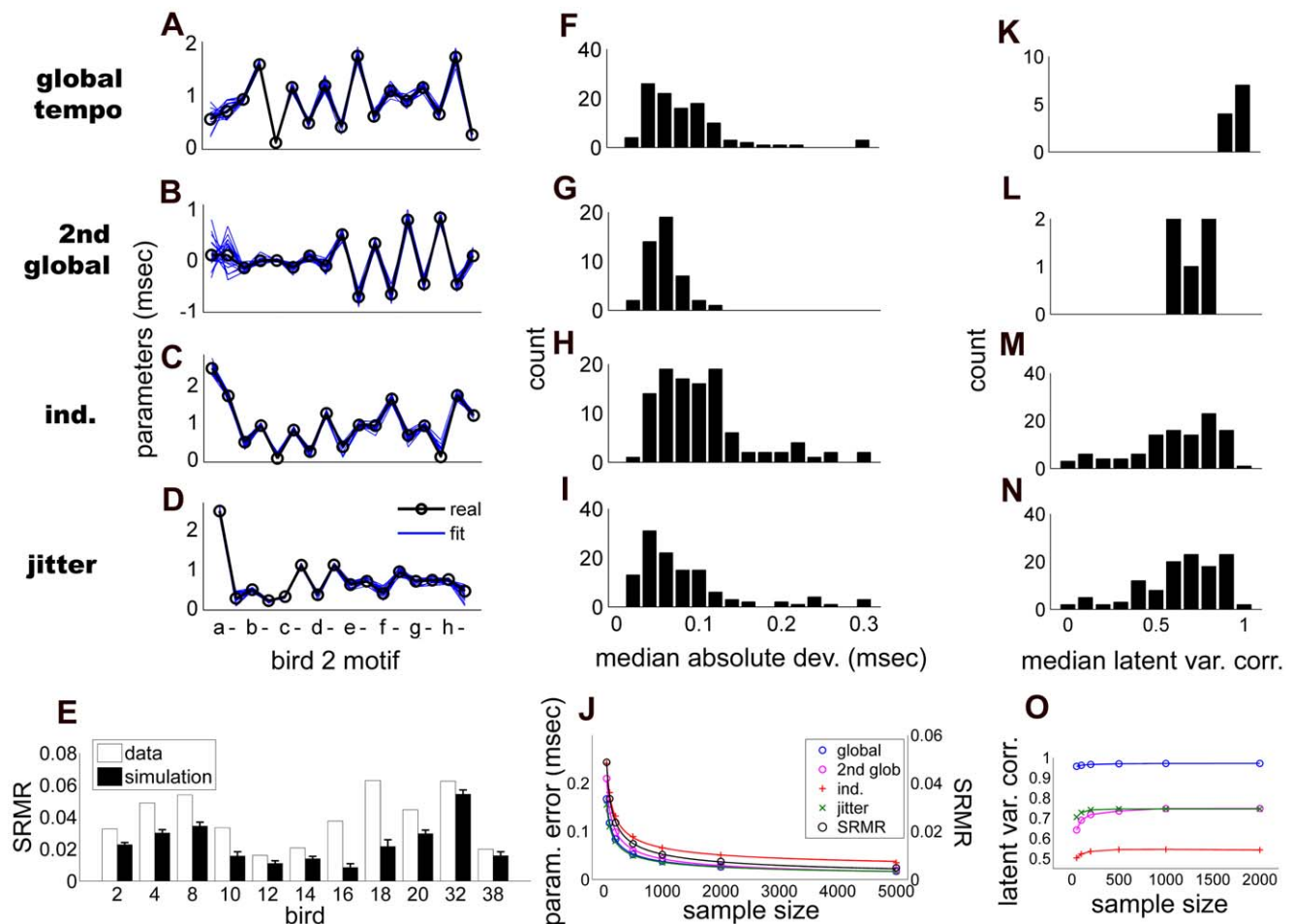


Figure 6. Results of Monte Carlo simulations based on real parameter distributions in the song data. For plots A-D, J and K-N, parameter type is designated by labels on the far left. A-D, Real parameter values (black lines), along with parameter estimates from 20 randomly selected simulations (blue) for one bird. E, Median SRMR across simulations for each bird (black bars, errorbars indicate median absolute deviation) vs. SRMR values from the real data (white bars). F-I, Median absolute deviation in parameter estimates from real parameters across all birds. J, Parameter error and SRMR as a function of sample size for simulations selected from 3 birds. Here, parameter error was taken as the median MAD across simulations and birds. Solid lines represent a regression of error and SRMR with the square root of sample size. K-N, Distribution of Pearson's correlation between the real and simulated latent variables. O, Median latent variable correlation from the same 3 birds taken as a function of sample size. Color legend is the same as in panel J. doi:10.1371/journal.pone.0037616.g006

showed a qualitatively close fit (Fig. 6J). We also computed the dependency of SRMR on sample size and found a similar relationship. Overall, deviation between estimated and known parameters tended to fall below 10 percent as song sample size increased above 200–500.

We then examined the accuracy of trial-by-trial maximum likelihood estimates of latent variables by computing Pearson's correlation between the real and simulated variables and taking the median across simulations (Fig. 6K–N). Global tempo showed remarkably strong correlations between 0.879 and 0.978 by bird (median $r=0.953$), while the 2nd global factor showed a weaker correlation between 0.610 and 0.786 (median $r=0.740$). Across all time intervals, independent and jitter latent variables showed correlations of 0.661 ± 0.156 and 0.684 ± 0.150 . Interestingly, correlations were only weakly dependent on sample size and appeared to reach a ceiling by ~ 500 – 1000 samples (Fig. 6O).

Finally, we tested the sensitivity of the model to the assumption that the underlying distributions of latent variables were Gaussian. To do this we reran simulations the same way as above (using actual sample sizes), except that for each simulation, and each latent variable, we randomly chose among a set of distributions that included a Gaussian and two Weibull distributions [30], one with significant negative skew ($\alpha=10$, $\beta=2$), and the other with positive skew ($\alpha=2$, $\beta=10$). Weibull distributions were mean-subtracted and scaled to have the same variance specified by the model. Interestingly, model fits and parameter error were similar to error when all distributions were Gaussian. SRMR ranged from 0.009 to 0.054 (median 0.022). Tempo parameter error was 0.065 ± 0.037 msec, while error in the 2nd global parameter was a 0.054 ± 0.0149 msec; independent parameter error, 0.085 ± 0.040 msec; and jitter error, 0.054 ± 0.051 msec. Correlations with the latent variables were also similar, with a median global tempo correlation of 0.954 ± 0.019 , while the 2nd global correlation median was 0.744 ± 0.066 , independent, 0.655 ± 0.201 and jitter 0.677 ± 0.179 . We next tested a more extreme violation of the Gaussian assumption by assigning exponential distributions to all latent variables (mean-subtracted and scaled as above). Even in this case model error was similar to when all distributions were Gaussian, with SRMR between 0.008 and 0.053 (median 0.021); respective errors among the tempo, second global, independent and jitter parameters of 0.065 ± 0.037 , 0.054 ± 0.015 , 0.080 ± 0.039 and 0.053 ± 0.048 msec; and respective correlations with latent variables of 0.954 ± 0.018 , 0.743 ± 0.065 , 0.663 ± 0.201 and 0.679 ± 0.179 .

Thus, the simulations indicated that the model was reasonably accurate at estimating parameters and trial-to-trial values of the latent variables, and robust to violations of the Gaussian assumption.

Discussion

We have presented a statistical model of action sequences that separates trial-to-trial variability in global timing, uncorrelated variation in the duration of individual sequence elements, and timing jitter at the boundaries between elements. We derived an algorithm based on expectation-maximization to find maximum likelihood estimates of both model parameters and trial-to-trial values of each latent factor. Applying the model to the songs of 11 adult male zebra finches, we found that the model provided very good fits to trial-to-trial variability in the durations of song syllables and the gaps of silence between them, with each of the three factors contributing roughly equal amounts of variability to individual song elements. In contrast, measurements of motif

length were dominated by the global factor since such variability accumulates in a correlated manner across song elements. Given that the neural mechanisms driving global variability are likely to be distinct from those driving both independent and jitter factors, experiments designed to evaluate the effect of behavioral or pharmacological manipulations of the system on timing variability may reach different conclusion based on whether timing variability is evaluated at the level of individual syllables or entire motifs [8].

Our timing variability model can be seen as an extension of the Wing-Kristofferson [16] model of timing in humans. The latter model has been used to investigate motor variability in studies where human subjects produced isochronous timing intervals via finger taps or other basic gestures [16,25–28,31,32]. The key insight of Wing and Kristofferson [16] was to exploit the fact that variability in event timing that is introduced downstream of the motor pattern generator necessarily introduces negative covariance between the intervals surrounding the event. This insight was used to separate the measured timing variance into a component that was attributed to the central clock, and motor variability introduced at the motor periphery. In general, clock variability scales with the duration of the interval, consistent with a process of accumulating variability of the time course of the interval, while motor variability shows little or no change with duration, consistent with variability introduced in a specific event that does not accumulate over time [16]. Consistent with these ideas, we were able to use negative covariance to segregate out a component of timing variance in zebra finch song that did not scale with interval duration. In contrast, the global and independent timing variability, hypothesized to relate to central circuits for song generation, did scale. Importantly, previous extensions of the Wing-Kristofferson model have also included a “tempo” variable, in which the speed of tapping slowly drifts over the course of an experiment [31–33].

In contrast to the original Wing-Kristofferson model and its extensions, the timing variability model yields data on heterogeneous action sequences and can measure how variability in each latent factor is spread across different sequence elements. Indeed, we found prominent and systematic differences in the variability of syllables and inter-syllable gaps. We also found that the gaps between motifs were more variable than within-motif gaps, and that this enhanced variability is expressed only in the independent component of timing variability, not in its sensitivity to global timing. Given the millisecond scale linkage between the timing of neural activity and song behavior [9–11], these data may be used to constrain mechanistic models of song production.

These findings may bear on how precise patterns of neural activity remain synchronized over disparate song nuclei. While several studies have suggested that HVC controls the timing of all neural activity [11,13,34,35], both left and right HVC appear to make independent contributions to song tempo [13]. Moreover, lesions to LMAN, the output of a basal ganglia circuit, decrease variability in syllable duration [8], suggesting that this nucleus contributes to timing variability as well. One idea for how areas remain in temporal alignment comes from electrophysiological evidence that left and right HVC are synchronized at specific time points within the song, and generally near syllable onsets [36]. Thus, while each HVC could run at a slightly different tempo, they may be periodically reset to a global tempo. If such resetting occurs preferentially at syllable onsets, this in turn would lead to a tradeoff between the durations of syllables and gaps, with the magnitude of the tradeoff depending on the mismatch between HVC and global tempo. If this mismatch were to undergo slow changes over time, then the tradeoff between syllable and gap lengths would be quite similar to the properties of the second

global factor in our behavioral data: it would not scale with syllable length but would show circadian modulation and slow drift. It may be the case that the tempo mismatch is smaller in some birds than others, explaining why the second factor reached significance in just 5 of 11 birds. This notion of a reset is consistent with recent recordings of air sac pressure during gaps, in which ongoing patterns of negative pressure are rapidly terminated near syllable onsets [12]. Finally, there is electrophysiological evidence for a separate source of signals that triggers the onset of motifs [29], which would be consistent with the especially large independent timing variability we find in inter-motif gaps because that source could introduce additional variability to the onset time of the first motif syllable.

In using the EM algorithm to separate the three hypothesized components of timing variance, the timing variability model also yields maximum likelihood estimates of the latent variables on a trial-to-trial basis. These values can then be related to variables that influence song timing but were not included in the original model. Here we have used this approach to quantify slow drift in timing that occurs on timescales much longer than between the production of individual songs. It is important to note that the original data used for this model consisted of songs collected in a variety of behavioral conditions (males were either alone, or singing in the presence of another adult or juvenile male). Both drift, as well as the magnitude of variability components, could conceivably be affected by this factor; future investigations could either track latent variables across conditions or estimate parameters separately for each. Regardless, timing drift has been previously reported in zebra finch song, and was shown to be highly correlated with drift in the duration of interburst intervals in the song nucleus RA [10]. That study did not decompose variability into different sources; it would be interesting to examine whether the tight correlation specifically stems from any of the individual components we have presented here.

Our use of EM for timing analysis falls within a broader theoretical effort in fitting probabilistic generative models to data [17,37]. This work suggests several extensions to the current timing variability model, including the use of Gaussian process or Kalman filters to explicitly model slower trial-by-trial drift in timing [37,38]. Another advantage of the generative model approach is that Monte Carlo simulations can be used to understand the reliability and limitations of the underlying model. Here we have examined the statistical reliability of the model by generating artificial data sets and then fitting the model to these data. Further insight into the robustness of the model can be obtained by altering key model assumptions, generating artificial data sets with the altered model, and then examining how well the original model fits these altered data. We have used this approach to argue that the model is robust to alterations in the assumption that the shape of the latent variable distribution is Gaussian. Further such investigations are beyond the scope of the present paper.

Behavioral and neural variability are often affected by multiple sources whose effects are shared by many different areas of the given system. Generative models address this problem by assigning latent variables to these different sources, and thus have the potential to help quantify the activity of hypothesized synergies in sensorimotor systems [39–41]. For precisely timed tasks such as zebra finch song, the coordination of activity throughout the system is a challenge, and the examination of timing variability may be particularly valuable in understanding the functional organization of the song circuit. By applying the methods developed here to a joint analysis of neural and behavioral timing,

future studies may yield a more complete understanding of the mechanisms linking neural activity and song behavior.

Methods

Ethics Statement

All bird care and housing was approved by the institutional animal care and use committee at the University of Maryland, College Park.

Model and Derivation

We fit the timing variability model using an EM algorithm similar to what has been used for maximum likelihood factor analysis [17,23,42]. We begin by rewriting equations 2 and 3 for the row-vector of mean-subtracted interval durations for sample n , and the resulting model covariance matrix \mathbf{S} :

$$\mathbf{x}_n = \mathbf{W}\mathbf{z}_n + \mathbf{D}\mathbf{u}_n + \eta_n \tag{4}$$

$$\mathbf{S} = \mathbf{W}\mathbf{W}^T + \mathbf{D}\mathbf{Q}\mathbf{D}^T + \mathbf{\Psi} \tag{5}$$

where \mathbf{D} is the differencing matrix described in Results. We will use $\tilde{\mathbf{S}}$ to denote the data covariance matrix, in order to distinguish it from \mathbf{S} .

The free parameters of the model are the global weights \mathbf{W} and the expected jitter and independent variances along the diagonals of \mathbf{Q} and $\mathbf{\Psi}$. We use vector θ to refer to the entire set of these parameters. For K time intervals in a sequence, $K-1$ modeled boundaries between intervals, and Q global variables, the dimensions of \mathbf{W} are $K \times Q$; \mathbf{Q} , $(K-1) \times (K-1)$; and $\mathbf{\Psi}$, $K \times K$. The total number of parameters in the model is thus $(Q+2)K-1$. In the original formulation of the model we assumed a single global dimension representing tempo, so in that case $Q=1$ and the total number of parameters is $3K-1$.

For a given parameter set, the model can be run in generative mode in which latent variables \mathbf{z}_n , \mathbf{u}_n , and η_n are generated and then combined to according to equation (4) to produce mean-subtracted interval durations \mathbf{x}_n . Our goal is to find a set of parameters θ that maximizes the likelihood of generating the set of interval durations that were actually observed, *i.e.* we want to maximize $p(\mathbf{x}|\theta)$. Although there is no closed-form solution to this problem in general, EM provides an iterative procedure for changing parameters so that each step of the algorithm increases the probability of generating the real data, *i.e.* $p(\mathbf{x}|\theta^{new}) \geq p(\mathbf{x}|\theta^{old})$.

Our derivation considers the independent variables η_n separately from the global and jitter variables \mathbf{z}_n and \mathbf{u}_n , and concatenating the latter two into a single $(K+Q-1) \times 1$ row vector \mathbf{v}_n . We also let \mathbf{A} denote the $K \times (K+Q-1)$ matrix obtained by concatenating \mathbf{W} and \mathbf{D} and let $\mathbf{\Phi}$ denote the $(K+Q-1) \times (K+Q-1)$ expected covariance matrix for \mathbf{v} :

$$\mathbf{v}_n = \begin{bmatrix} \mathbf{z}_n \\ \mathbf{u}_n \end{bmatrix}, \quad \mathbf{A} = [\mathbf{W} \quad \mathbf{D}], \quad \mathbf{\Phi} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{Q} \end{bmatrix} \tag{6}$$

where \mathbf{I} is the $Q \times Q$ identity matrix, representing the assumed unit covariance of \mathbf{z}_n , and $\mathbf{0}$ is a $Q \times (K-1)$ matrix of zeroes. So $\mathbf{A}\mathbf{v}_n = \mathbf{W}\mathbf{z}_n + \mathbf{D}\mathbf{u}_n$ and

$$\mathbf{x}_n = \mathbf{A}\mathbf{v}_n + \eta_n \tag{7}$$

$$\mathbf{S} = \mathbf{A}\mathbf{\Phi}\mathbf{A} + \eta \tag{8}$$

Equation 8 may be viewed as a kind of ‘‘confirmatory factor analysis’’ or ‘‘covariance structure analysis’’ [18,19,21,22], which are generalizations of factor analysis in which \mathbf{A} and $\mathbf{\Phi}$ contain a combination of free and fixed parameters.

The basic idea in the EM algorithm stems from obtaining the probability we seek to maximize, $p(\mathbf{x}|\theta)$, by marginalizing over the latent variables \mathbf{v} , i.e. $p(\mathbf{x}|\theta) = \int p(\mathbf{x}, \mathbf{v}|\theta)p(\mathbf{v}|\mathbf{x}, \theta)\delta\mathbf{v}$. Instead of dealing with both terms in this integral at the same time, EM is an iterative algorithm that addresses each term separately. In the ‘E’ step, one fixes the parameters and finds the probability distribution (q) of latent variables in each sample at the current values of the parameters, $q(\mathbf{v}_n) = p(\mathbf{v}_n|\mathbf{x}_n, \theta)$. The structure of the model implies that the conditional distribution of the data $p(\mathbf{x}_n|\mathbf{v}_n, \theta)$ is a Gaussian with mean $\mathbf{A}\mathbf{v}_n$ and covariance matrix equal to the independent covariance matrix $\mathbf{\Psi}$. From Bayes’ Theorem for Gaussian variables [17], it follows that $q(\mathbf{v}_n)$ is also a Gaussian with covariance defined by

$$\mathbf{G} = (\mathbf{\Phi}^{-1} + \mathbf{A}^T \eta^{-1} \mathbf{A})^{-1} \tag{9}$$

and expected mean

$$\mathcal{E}\{\mathbf{v}_n\} = \mathbf{G}\mathbf{A}^T \mathbf{\Psi}^{-1} \mathbf{x}_n \tag{10}$$

where we use the notation $\mathcal{E}\{\mathbf{y}\}$ to denote the expected value of any given variable \mathbf{y} under the distribution $q(\mathbf{v})$. We also require the 2nd moment of $q(\mathbf{v}_n)$, which is given by

$$\mathcal{E}\{\mathbf{v}_n \mathbf{v}_n^T\} = \mathcal{E}\{\mathbf{v}_n\} \mathcal{E}\{\mathbf{v}_n\}^T + \mathbf{G} \tag{11}$$

using the basic result that $\mathcal{E}\{\mathbf{y}\mathbf{y}^T\} = \mathcal{E}\{\mathbf{y}\} \mathcal{E}\{\mathbf{y}\}^T + cov(\mathbf{y})$.

In order to analyze the global and jitter variables separately we can simply separate out the expected values from $\mathcal{E}\{\mathbf{v}_n\} = [\mathcal{E}\{\mathbf{z}_n\} \quad \mathcal{E}\{\mathbf{u}_n\}]$. Below we will also make use of the 2nd moments of the original latent variables and use the fact that

$$\mathcal{E}\{\mathbf{v}_n \mathbf{v}_n^T\} = \begin{bmatrix} \mathcal{E}\{\mathbf{z}_n \mathbf{z}_n^T\} & \mathcal{E}\{\mathbf{z}_n \mathbf{u}_n^T\} \\ \mathcal{E}\{\mathbf{u}_n \mathbf{z}_n^T\} & \mathcal{E}\{\mathbf{u}_n \mathbf{u}_n^T\} \end{bmatrix} \tag{12}$$

It is important to note that while the latent variables are independent of each other by definition, they are not necessarily independent under the posterior distribution $q(\mathbf{v}_n)$. For example, $\mathcal{E}\{\mathbf{z}_n \mathbf{u}_n^T\} \neq 0$. Thus, while we seek to separate the global and jitter parameters as well as corresponding latent variables, the E step is facilitated by their concatenation.

In the ‘M’ step, one finds new values for \mathbf{q} that maximize the expected log likelihood of the joint distribution of the data and the latent variables under $q(\mathbf{v}_n)$. That is, we want to find θ^{new} that maximizes

$$\begin{aligned} \mathcal{L} &= \mathcal{E}\{\log(\prod_n p(\mathbf{x}_n, \mathbf{v}_n|\theta))\} \\ &= \sum_n \mathcal{E}\{\log(p(\mathbf{x}_n|\mathbf{v}_n, \theta))\} + \mathcal{E}\{\log(p(\mathbf{v}_n|\theta))\} \end{aligned} \tag{13}$$

Using the standard log likelihood function of a multivariate Gaussian distribution, along with our definition of $p(\mathbf{x}_n|\mathbf{v}_n, \theta)$ as a Gaussian with mean $\mathbf{A}\mathbf{v}_n$ and covariance matrix $\mathbf{\Psi}$,

$$\begin{aligned} \sum_n \log(p(\mathbf{x}_n|\mathbf{v}_n, \theta)) &= -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\det \mathbf{\Psi}) \\ &\quad - \frac{1}{2} \sum_n (\mathbf{x}_n - \mathbf{A}\mathbf{v}_n)^T \mathbf{\Psi}^{-1} (\mathbf{x}_n - \mathbf{A}\mathbf{v}_n) \end{aligned} \tag{14}$$

Conditioned solely on the parameters θ , the distribution of \mathbf{v} is also Gaussian with covariance matrix $\mathbf{\Phi}$. So

$$\begin{aligned} \sum_n \log(p(\mathbf{v}_n|\theta)) &= -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\det \mathbf{\Phi}) \\ &\quad - \frac{1}{2} \sum_n \mathbf{v}_n^T \mathbf{\Phi}^{-1} \mathbf{v}_n \end{aligned} \tag{15}$$

We seek values of \mathbf{W} , $\mathbf{\Omega}$, and $\mathbf{\Psi}$ that maximize \mathcal{L} . Thus, for each variable, we take the partial derivative of \mathcal{L} with respect to the variable, set the derivative equal to zero and solve for the variable. Below we will make use of the fact that taking expected values is a linear operation, so the derivative of the expected value is equal to the expected value of the derivative.

We begin with $\mathbf{\Omega}$, noting that the variable is contained in $\mathbf{\Phi}$ so the partial derivative only depends on the last two terms of equation 15. Since $\mathbf{\Phi}$ is simply a concatenation that includes diagonal matrices \mathbf{I} and $\mathbf{\Omega}$ as in equation 6, we can make use of the fact that $\det \mathbf{\Phi} = (\det \mathbf{I})(\det \mathbf{\Omega}) = \det \mathbf{\Omega}$. Also, $\mathbf{v}_n^T \mathbf{\Phi}^{-1} \mathbf{v}_n = \mathbf{z}_n^T \mathbf{z}_n + \mathbf{u}_n^T \mathbf{\Omega}^{-1} \mathbf{u}_n$, and we can discard the term with \mathbf{z} for the partial derivative. Therefore,

$$\begin{aligned} \partial \mathcal{L} / \partial \mathbf{\Omega} &= \partial / \partial \mathbf{\Omega} \left[\mathcal{E}\left\{-\frac{N}{2} \log(\det \mathbf{\Omega}) - \frac{1}{2} \sum_n \mathbf{u}_n^T \mathbf{\Omega}^{-1} \mathbf{u}_n\right\}\right] \\ &= -\frac{N}{2} \mathbf{\Omega}^{-1} + \frac{1}{2} \sum_n \mathbf{\Omega}^{-1} \mathcal{E}\{\mathbf{u}_n \mathbf{u}_n^T\} \mathbf{\Omega}^{-1} \end{aligned} \tag{16}$$

which makes use of the identity $\partial / \partial \mathbf{B} [\mathbf{y}^T \mathbf{B}^{-1} \mathbf{y}] = -\mathbf{B}^{-1} \mathbf{y} \mathbf{y}^T \mathbf{B}^{-1}$ when \mathbf{B} is symmetric. Setting the derivative to zero and solving for $\mathbf{\Omega}$ we have

$$\mathbf{\Omega}^{new} = \frac{1}{N} \sum_n \mathcal{E}\{\mathbf{u}_n \mathbf{u}_n^T\} \tag{17}$$

This is simply the standard solution for the expected 2nd moment of any Gaussian distribution (in this case $p(\mathbf{v}_n|\theta)$) and follows directly from the fact that the partial derivative does not depend on any terms in $p(\mathbf{x}_n|\mathbf{v}_n, \theta)$.

To maximize \mathcal{L} with respect to \mathbf{W} we first note that the variable is contained in \mathbf{A} , so the only term that is relevant to the partial derivative is the quadratic in equation 14, which can be rewritten as $(\mathbf{x}_n - \mathbf{A}\mathbf{v}_n)^T \mathbf{\Psi}^{-1} (\mathbf{x}_n - \mathbf{A}\mathbf{v}_n) = (\mathbf{W}\mathbf{z}_n + \mathbf{D}\mathbf{u}_n - \mathbf{x}_n)^T \mathbf{\Psi}^{-1} (\mathbf{W}\mathbf{z}_n + \mathbf{D}\mathbf{u}_n - \mathbf{x}_n)$. We also make use of the identity $\partial / \partial \mathbf{B} [(\mathbf{B}\mathbf{y} + \mathbf{w})^T \mathbf{C}(\mathbf{B}\mathbf{y} + \mathbf{w})] = 2\mathbf{C}(\mathbf{B}\mathbf{y} + \mathbf{w})\mathbf{y}^T$, to give

$$\partial \mathcal{L} / \partial \mathbf{W} = \mathcal{E}\{2\mathbf{\Psi}^{-1} (\mathbf{W}\mathbf{z}_n + \mathbf{D}\mathbf{u}_n - \mathbf{x}_n) \mathbf{z}_n^T\} \tag{18}$$

Setting the derivative to 0 and solving for \mathbf{W} , we have

$$\mathbf{W}^{new} = \left[\sum_n \mathbf{x}_n \mathcal{E}\{\mathbf{z}_n\}^T - \mathbf{D} \mathcal{E}\{\mathbf{u}_n \mathbf{z}_n^T\} \right] \left[\sum_n \mathcal{E}\{\mathbf{z}_n \mathbf{z}_n^T\} \right]^{-1} \tag{19}$$

To maximize \mathcal{L} with respect to Ψ we require the first two terms of equation 15 and use \mathbf{W}^{new} as well as the identity used to derive \mathbf{W}^{new} , finding:

$$\partial\mathcal{L}/\partial\Psi = \mathcal{E}\left\{-\frac{N}{2}\Psi^{-1} + \sum_n \Psi^{-1}(\mathbf{x}_n - \mathbf{A}^{\text{new}}\mathbf{v}_n)(\mathbf{x}_n - \mathbf{A}^{\text{new}}\mathbf{v}_n)^T\Psi^{-1}\right\} \quad (20)$$

where $\mathbf{A}^{\text{new}} = [\mathbf{W}^{\text{new}}\mathbf{D}]$.

Setting the expected value of the partial derivative to zero and solving for Ψ :

$$\Psi^{\text{new}} = \text{diag}\left\{\tilde{\mathbf{S}} + \frac{1}{N}\sum_n \mathbf{A}^{\text{new}}\mathcal{E}\{\mathbf{v}_n\mathbf{v}_n^T\}(\mathbf{A}^{\text{new}})^T - 2\mathbf{A}^{\text{new}}\mathbf{x}_n\mathcal{E}\{\mathbf{v}_n^T\}\right\} \quad (21)$$

where we have used $\tilde{\mathbf{S}} = \frac{1}{N}\sum_n \mathbf{x}_n\mathbf{x}_n^T$ and “diag” is the operation that creates a diagonal matrix by setting all off-diagonal elements to zero.

Model Implementation

We implemented the model with custom code written for Matlab (Mathworks, Natick, MA). Like all “hill-climbing” algorithms, EM is subject to getting stuck in a local maximum. Therefore, for each optimization we ran EM 100 times starting from different sets of initial parameter values, chosen from a uniform distribution ranging from zero to a maximum value determined by the variability of the corresponding interval. For the global weight \mathbf{W}_k initial values ranged from 0 to the standard deviation of \mathbf{x}_k . Initial values of the independent variance for interval k ranged from 0 to the sample variance of that interval. Since a jitter parameters are not associated with unique intervals and overall contribute twice as much variance to intervals as other parameters, initial values for the jitter variances ranged from 0 to half of average interval variance. For each set of initial conditions, we stopped the algorithm when it satisfied the relatively loose convergence criterion that the log-likelihood of the data given the parameters (L) increase by $<0.1\%$ over a given iteration. Here, $L = \sum_n \log(p(\mathbf{x}_n|\theta))$, which is a Gaussian with a mean of 0 and covariance \mathbf{S} . Using the standard formula for the likelihood function of a Gaussian and the fact that $\tilde{\mathbf{S}} = \frac{1}{N}\sum_n \mathbf{x}_n\mathbf{x}_n^T$, it is easily verified that

$$L = -N/2[D\log(2\pi) + \log(\det\tilde{\mathbf{S}}) + \text{Tr}((\tilde{\mathbf{S}}\mathbf{S})^{-1})] \quad (22)$$

After running the algorithm from 100 initial conditions we picked the 5 parameter sets with the highest log-likelihoods and continued running the algorithm on those parameters until the global and timing jitter parameters changed by $<0.01\%$ from the previous iteration. We omitted independent variability because we expected that parameter set to converge more quickly; we have verified that assumption and note that there is no significant difference in parameter estimates or error in Monte Carlo simulations when independent parameters are directly required to converge in the code. After parameters converged, we then picked the parameter set associated with the highest log-likelihood under that criterion. Across song data and Monte Carlo simulations the algorithm never failed to converge within a

10,000 iteration limit specified in the computer code. The number of iterations required to meet the looser convergence criterion typically ranged from $\sim 50-150$, while for the stricter criterion the range was $\sim 300-3000$ with the large majority being <1000 .

Model fit evaluation. Model fits were evaluated using the standardized Root Mean Squared Residual (SRMR) between the data and model correlation matrices [24]. We let the residual correlation \mathbf{R}_{jk} for time intervals j and k equal the difference between Pearson’s correlation coefficient for j and k determined from the data, minus the covariance of these intervals predicted by the modeled normalized by the standard deviations in the data (σ_j and σ_k):

$$\mathbf{R}_{jk} = \frac{\tilde{\mathbf{S}}_{jk}}{\sigma_j\sigma_k} - \frac{\mathbf{S}_{jk}}{\sigma_j\sigma_k} \quad (23)$$

SRMR is just the root mean square average of \mathbf{R}_{jk} , where we only consider values with $j \geq k$ since \mathbf{R} is symmetric:

$$\text{SRMR} = \left[\left(\sum_{j \geq k} \mathbf{R}_{jk}^2\right)/(K(K+1)/2)\right]^{\frac{1}{2}} \quad (24)$$

Additional global factors. Initially, all model fits were performed with only 1 global timing factor, *i.e.* \mathbf{z} is one dimensional. However, the song data yielded a few SRMR values that we considered relatively large, and a qualitative examination suggested that significant structure in the off-diagonal elements of the data covariance matrix was not being captured by the model. We thus considered additional global timing factors by increasing the number of dimensions (Q) in \mathbf{z} . Here, each dimension of \mathbf{z} remains unit-Gaussian and is independent of the other dimensions, so the expected covariance of \mathbf{z} is the $Q \times Q$ identity matrix. The derivation of the EM algorithm is unchanged; \mathbf{W} is simply a $K \times Q$ dimensional matrix rather than a $K \times 1$ dimensional column vector.

To determine the optimal number of global factors per bird, we ran the entire EM algorithm (including 100 initial conditions) for each value of Q ranging from 1–4. We then computed the Bayesian Information Criterion (BIC) [43], and chose the dimensionality associated with the lowest BIC. Specifically, we calculated BIC as $-2L + M \log(N)$ where L is the log-likelihood defined in equation 22, M is the total number of model parameters $= (Q+2)K - 1$ and N is the number of samples as before.

For a number of birds the BIC analysis indicated that >1 global timing factor provided the best fit to the data. However, the columns of \mathbf{W} , and hence the global factor weights are not uniquely specified by the algorithm: multiple estimates of \mathbf{W} can yield the same global covariance matrix $\mathbf{W}\mathbf{W}^T$. This interpretive difficulty is common in factor analysis and is typically addressed using various transformations, called factor rotations, that are based on a priori assumptions about the relationship among latent variables. Here, we consider an orthonormal transformation (\mathbf{U}) in latent variable space. In this new basis, the transformation from latent to observable space is given by the matrix $\mathbf{V} = \mathbf{W}\mathbf{U}$, so $\mathbf{V}\mathbf{V}^T = \mathbf{W}\mathbf{U}\mathbf{U}^T\mathbf{W}^T = \mathbf{W}\mathbf{W}^T$.

We chose a transformation \mathbf{U} such that all the global variability in total sequence length was captured by the first dimension in latent variable space. For any latent vector \mathbf{y} , the global contribution to total sequence length can be determined by taking the inner product of the image of \mathbf{y} under \mathbf{W} with the D

dimensional column vector of ones $[1,1,1\dots]^T$, which we designate as $\mathbf{1}$, *i.e.* global sequence length is $\mathbf{1}^T \mathbf{W} \mathbf{y}$. Thus, all global variation in sequence length is determined by variation in the direction of the unit-length latent vector

$$\mathbf{y}_1 = \mathbf{W}^T \mathbf{1} / \|\mathbf{W}^T \mathbf{1}\| \quad (25)$$

Importantly, any other direction \mathbf{y}_r in latent variable space that is orthogonal to \mathbf{y}_1 will yield no variation in global sequence length: by definition of \mathbf{y}_1 , this is equal to $\|\mathbf{W}^T \mathbf{1}\| \mathbf{y}_1^T \mathbf{y}_r = 0$. Therefore, a rotation matrix \mathbf{U} whose first column is given by \mathbf{y}_1 yields a transformation where all global sequence length variation is confined to the first factor, and the first column of \mathbf{V} should be the image of \mathbf{y}_1 under the transformation \mathbf{W} : $\mathbf{V}_1 = \mathbf{W} \mathbf{y}_1 = \mathbf{W} \mathbf{W}^T \mathbf{1} / \|\mathbf{W}^T \mathbf{1}\|$.

For $Q=2$ global factors, we need only to complete the basis by setting the second column of \mathbf{U} to be a unit vector in the unique direction perpendicular to \mathbf{y}_1 . However, for the one bird with $Q=3$, there were an infinite number of solutions since there is no unique orthonormal basis to the remaining >1 dimensional latent space, so we sought a more general solution. Thus, to fill out the remaining columns of \mathbf{V} , we subtracted off the variations in the direction of \mathbf{V}_1 by calculating the covariance matrix $\mathbf{W} \mathbf{W}^T - \mathbf{V}_1 \mathbf{V}_1^T$. We then performed a principal components analysis on this covariance matrix by computing $\mathbf{\Lambda}$, the $(Q-1) \times (Q-1)$ diagonal matrix of non-zero eigenvalues, and \mathbf{P} , the $K \times (Q-1)$ matrix of associated eigenvectors. Then we set

$$\mathbf{V} = [\mathbf{V}_1 \quad \mathbf{V}_r] \quad (26)$$

where

$$\mathbf{V}_r = \mathbf{P} \mathbf{\Lambda}^{1/2} \quad (27)$$

By convention we can sort the columns of \mathbf{V}_r from largest to smallest eigenvalue. Thus, when comparing the 2nd columns of \mathbf{V}_r across birds we are effectively choosing the dimensions that maximize the sum of the global data variances that remain after subtracting out global sequence length variability. Note that by construction, $\mathbf{V}_r \mathbf{V}_r^T = \mathbf{W} \mathbf{W}^T - \mathbf{V}_1 \mathbf{V}_1^T$, ensuring that $\mathbf{V} \mathbf{V}^T = \mathbf{W} \mathbf{W}^T$.

Song Data

We used song data from 11 adult zebra finch males >400 days-post hatch. All care and housing was approved by the institutional animal care and use committee at the University of Maryland, College Park. Our basic procedure for collecting and extracting song data was based on custom code written in Matlab and has been previously reported in [14,15]; we provide a brief description below.

Recordings were made from sound isolation chambers (Industrial Acoustics, Bronx, NY), which contained two cages separated by 18 cm and two directional microphones (Pro 45; Audio-Technica, Stow, Ohio). Signals were digitized at 24,414.1 Hz, and

ongoing data were selected using a circular buffer and a sliding window amplitude algorithm. ‘‘Sound clips’’ separated by <200 msec were included in the same ‘‘recording’’ and clip onset times were indicated by filling the gaps between clips with zeros.

For each bird, we gathered an initial random sample of 1000 recordings >2 sec long and had maximum power from the side on which the target bird was housed. We analyzed recordings using the log-amplitude of the fast-Fourier transform (FFT) with a 256-point (10.49 msec) window moved forward in 128-point steps and excluded frequencies outside the 1.7–9 kHz range from all subsequent analysis because song structure is less reliable outside that range. We then used an automated template-matching algorithm [14] to identify individual song syllables, and selected out recordings which contained repeated sequences of the most commonly produced motif. Our final sample of songs ranged from 215–885 per bird.

After we identified syllables and syllable sequences we measured the onset and offset times of each syllable with a more fine-grained algorithm: First, we recalculated spectrograms from the original signal using FFTs with a 128-point window slid forward in 4-point steps, yielding 0.16 msec time bins. Although previous research [14,15] had been based on log-amplitudes, for this study we used raw amplitudes, which we have found to be more reliable for timing measurements. We then computed time-derivative spectrograms as differences in amplitude in time-adjacent bins. We next smoothed in time the resulting spectrograms with a 64-point Gaussian window that had a 25.6-point (~ 5 msec) standard deviation. This windowing was found to be a good compromise between the competing demands of averaging out extraneous peaks and troughs in the time-derivative vs. preserving the peaks and troughs that we had defined as syllable onsets and offsets.

Finally, for each syllable we constructed a new template based on the time-derivative spectrogram and used a novel dynamic time-warping algorithm to determine the precise times at which onsets and offsets occurred in individual recordings of those syllables; see [14] especially for more detail on this process. Onsets/offsets were generally identified as salient time-derivative peaks/troughs in the template spectrogram; these correspond to inflection points in the rise and fall of energy at the beginning and ends of syllables.

Supporting Information

Matlab S1 A.zip file containing Matlab functions for model implementation. Also contains a.txt file giving an overview of the functions. (ZIP)

Acknowledgments

We thank Sarah Smith and Dr. Christina Castellino for helpful comments.

Author Contributions

Conceived and designed the experiments: CMG TWT. Performed the experiments: CMG TWT. Analyzed the data: CMG TWT. Contributed reagents/materials/analysis tools: CMG TWT. Wrote the paper: CMG TWT.

References

1. Faisal A, Selen L, Wolpert D (2008) Noise in the nervous system. *Nature Rev Neurosci* 9: 292–303.
2. Schoppik D, Nagel K, Lisberger S (2008) Cortical mechanisms of smooth eye movements revealed by dynamic covariations of neural and behavioral responses. *Neuron* 58: 248–260.

3. Churchland M, Afshar A, Shenoy K (2006) A central source of movement variability. *Neuron* 52: 1085–1096.
4. Sossinka R, Bohner J (1980) Song types in the zebra finch *poephila guttata castanotis*. *Z Tierpsychol* 53: 123–132.
5. Kao MH, Doupe AJ, Brainard MS (2005) Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature* 433: 638–43.
6. Kao MH, Brainard MS (2006) Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *J Neurophysiol* 96: 1441–1455.
7. Stepanek L, Doupe A (2010) Activity in a cortical-basal ganglia circuit for song is required for social context-dependent vocal variability. *J Neurophysiol* 104: 2474–86.
8. Thompson J, Basista M, Wu W, Bertram R, Johnson F (2011) Dual pre-motor contribution to songbird syllable variation. *J Neurosci* 31: 5967–09.
9. Yu AC, Margoliash D (1996) Temporal hierarchical control of singing in birds. *Science* 273: 1871–5.
10. Chi Z, Margoliash D (2001) Temporal precision and temporal drift in brain and behavior of zebra finch song. *Neuron* 32: 899–910.
11. Hahnloser RH, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419: 65–70.
12. Andalman A, Foerster J, Fee M (2011) Control of vocal and respiratory patterns in birdsong: Dissection of forebrain and brainstem mechanisms using temperature. *PLoS ONE* 6: e25461.
13. Long M, Fee M (2008) Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* 456: 189–194.
14. Glaze CM, Troyer TW (2007) Behavioral measurements of a temporally precise motor code for birdsong. *J Neurosci* 27: 7631–7639.
15. Glaze CM, Troyer TW (2006) Temporal structure in zebra finch song: Implications for motor coding. *J Neurosci* 26: 991–1005.
16. Wing A, Kristofferson A (1973) Response delays and the timing of discrete motor responses. *Perception & Psychophysics* 14: 5–12.
17. Bishop C (2006) *Pattern Recognition and Machine Learning*. New York, NY: Springer.
18. Bollen K (2002) Latent variables in psychology and the social sciences. *Ann Rev Psychol* 53: 605–634.
19. Bentler P, Dudgeon P (1996) Covariance structure analysis: statistical practice, theory, and directions. *Ann Rev Psychol* 47: 563–592.
20. Saar S, Mitra P (2008) A technique for characterizing the development of rhythms in bird song. *PLoS ONE* 3: e1461.
21. Jöreskog K (1978) Structural analysis of covariance and correlation matrices. *Psychometrika* 43: 443–477.
22. Jöreskog K (1969) A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* 34: 182–202.
23. Rubin D, Thayer D (1982) EM algorithms for ml factor analysis. *Psychometrika* 47: 69–76.
24. Hu L, Bentler P (1998) Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychol Methods* 3: 424–453.
25. Zarco W, Merchant H, Prado L, Mendez J (2009) Subsecond timing in primates: Comparison of interval production between human subjects and rhesus monkeys. *J Neurophysiol* 102: 3191–3202.
26. Ivry R, Hazeltine R (1995) Perception and production of temporal intervals across a range of durations- evidence for a common timing mechanism. *J Exp Psychol Human* 21: 3–18.
27. Heuer H (1988) Testing the invariance of relative timing: comment on Gentner (1987). *Psychol Rev* 95: 552–8.
28. Gentner DR (1987) Timing of skilled motor performance: Tests of the proportional duration model. *Psychol Rev* 94: 255–276.
29. Schmidt M, Ashmore R (2008) Integrating breathing and singing: forebrain and brainstem mechanisms. In: Zeigler H, Marler P, editors. *Neuroscience of Birdsong*. Cambridge: Cambridge University Press. 115–135.
30. Devore J (2004) *Probability and Statistics for Engineering and the Sciences*, 6th ed. Belmont, CA: Brooks/Cole.
31. Lemoine L, Delignières D (2009) Detrended windowed (lag one) autocorrelation: A new method for distinguishing between event-based and emergent timing. *Q J Exp Psycho* 62: 585–604.
32. Collier G, Ogden R (2004) Adding drift to the decomposition of simple isochronous tapping: An extension of the wingkristofferson model. *J Exp Psychol Hum Percept Perform* 30: 853–72.
33. Ogden R, Collier G (2002) Inference on variance components of autocorrelated sequences in the presence of drift. *J Nonparametr Stat* 14: 409–420.
34. Long M, Jin D, Fee M (2010) Support for a synaptic chain model of neuronal sequence generation. *Nature* 468: 394–399.
35. Fee MS, Kozhevnikov AA, Hahnloser RH (2004) Neural mechanisms of vocal sequence generation in the songbird. *Ann N Y Acad Sci* 1016: 153–70.
36. Schmidt MF (2003) Pattern of interhemispheric synchronization in hvc during singing correlates with key transitions in the song pattern. *J Neurophysiol* 90: 3931–49.
37. Roweis S, Ghahramani Z (1999) A unifying review of linear gaussian models. *Neural Comput* 11: 305–345.
38. Yu B, Cunningham J, Santhanam G, Ryu S, Shenoy K, et al. (2009) Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *J Neurophysiol* 102: 614–635.
39. Tresch M, Cheung V, d'Avella A (2006) Matrix factorization algorithms for the identification of muscle synergies: evaluation on simulated and experimental data sets. *J Neurophysiol* 95: 2199–2212.
40. Bizzi E, Cheung V, d'Avella A, Saltiel P, Tresch M (2008) Combining modules for movement. *Brain Res Rev* 57: 125–133.
41. Tresch M, Jarc A (2009) The case for and against muscle synergies. *Curr Opin Neurobiol* 19: 601–607.
42. Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the em algorithm. *J Roy Stat Soc B Met* 39: 1–38.
43. Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6: 461–464.