CrossMark

ORIGINAL PAPER

Variability in the temporal parameters in the song of the Bengalese finch (*Lonchura striata* var. *domestica*)

Ryosuke O. Tachibana[1,2] · **Takuya Koumura**[1,2] · **Kazuo Okanoya**[1,3]

Abstract Birdsong provides a unique model for studying the control mechanisms of complex sequential behaviors. The present study aimed to demonstrate that multiple factors affect temporal control in the song production. We analyzed the song of Bengalese finches in various time ranges to address factors that affected the duration of acoustic elements (notes) and silent intervals (gaps). The gaps showed more jitter across song renditions than did notes. Gaps had longer duration in branching points of song sequence than in stereotypic transitions, and the duration of a gap was correlated with the duration of the note that preceded the gap. When looking at the variation among song renditions, we found notable factors in three time ranges: within-day drift, within-bout changes, and local jitter. Note durations shortened over time from morning to evening. Within each song bout note durations lengthened as singing progressed, while gap durations lengthened only during the late part of song bout. Further analysis after removing these drift factors confirmed that the jitter remained in local song sequences. These results suggest distinct sources of temporal variability exist at multiple levels on the basis of this note–gap relationship, and that song comprised a mixture of these sources.

Introduction

Speech production and music performance require rapid and accurate control of sequential patterns of a variety of sound properties. For the study of a timing control mechanism of such complex motor sequences, songbirds provide a unique research model. Their songs consist of brief acoustic elements, and they appear to control the timing of these elements quite accurately. The present study aimed to assess in detail the tendencies for accuracy and variability of the song timing control in Bengalese finches (*Lonchura striata* var. *domestica*) and to understand the underlying neural mechanisms responsible for such tendencies.

The neural circuit in the song system that contributes to motor control consists of a descending pathway via a direct linkage from the premotor nucleus HVC (abbreviation used as proper name) to the robust nucleus of the arcopallium (RA). HVC has been reported to be involved in generating the timing and sequencing of song (Yu and Margoliash 1996; Hahnloser et al. 2002; Long and Fee 2008; Fee and Scharff 2010). Internal circuits within HVC produce regular timed sequences of neuronal bursts, and could be modeled as a synfire chain which can generate successive regular discharges by interlinked neurons (Hahnloser et al. 2002; Long et al. 2010). Similarly, another study argued that both timing and motor gestural

✉ Kazuo Okanoya
cokanoya@mail.ecc.u-tokyo.ac.jp

1   Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan

2   Research Fellow of Japan Society for the Promotion of Science, Tokyo, Japan

3   Cognition and Behavior Joint Laboratory, RIKEN Brain Science Institute, Saitama, Japan

information may be encoded in the spikes of HVC neurons (Amador et al. 2013). RA projects to motor neurons controlling the vocal muscles and respiratory system (Suthers and Margoliash 2002), and contributes to the control of song acoustical features (Leonardo and Fee 2005; Sober et al. 2008) by assigning actual motor commands to each time point of the HVC bursts (Fee and Scharff 2010). A mild cooling of HVC (to slow its bursting activity) resulted in slowing down the song tempo, whereas cooling of RA did not (Long and Fee 2008). These studies strongly suggest that HVC is at the top of the processing stream of song timing control.

A repetition-by-repetition variance of song acoustical properties has drawn much interest in the study of song learning and maintenance in finches (for example, Scharff and Nottebohm 1991; Kao et al. 2005; Tumer and Brainard 2007; Andalman and Fee 2009; Fee and Goldberg 2011; Sober and Brainard 2012; Woolley and Kao 2015). Birdsong also has variability in several types of temporal properties. Previous studies on the zebra finch, one model animal used in songbird research, have reported that a within-day, slow pattern of variation in song tempo was observed repeatedly over days (Glaze and Troyer 2006, 2012; Aronov and Fee 2012). This circadian pattern in song tempo appeared to depend on the temperature variation of the brain, specifically of the HVC nucleus (Aronov and Fee 2012). On the other hand, an analysis of local timing structures that separately assessed durations of acoustic elements (syllables or notes) and silent intervals (gaps) has suggested a different factor for this temporal variability. A given acoustic element showed a stronger duration trade-off relationship with its following gap, compared with its preceding gap (Glaze and Troyer 2006), suggesting that sound onsets (but not offsets) represent alignment cues for the song tempo. A greater variability seen in gaps than in syllables has been also reported and suggests involvement of different timing factors for each. (Glaze and Troyer 2006; Cooper and Goller 2006; Andalman et al. 2011; Glaze and Troyer 2012). Last, 'motif-boundary' gaps were relatively longer and more variable than 'within-motif' gaps (Glaze and Troyer 2006, 2012).

Recent advances in studies using Bengalese finches have provided an opportunity to assess the control mechanism of song sequencing with more variable, stochastic transition patterns (Okanoya 2004; Sakata and Brainard 2006; Warren et al. 2012). In Bengalese finch songs, sound intervals segmented by silent intervals tend to be shorter and have simpler acoustical patterns compared to zebra finch songs (and thus we term the sound elements in Bengalese finch songs 'notes' in this article). Further assessment of how temporal aspects of songs are similar between Bengalese and zebra finches would be beneficial for future neuroscientific studies of these two species.

Here, we assessed one-day recordings of Bengalese finch songs to study the variability of note and gap durations and determine how much temporal variation is characteristic of song. We analyzed song data based on note pairs. One previous study of zebra finches used an analysis method that integrated local and global temporal factors into one analysis model (Glaze and Troyer 2006, 2012). However, we did not employ this method in our analysis because Bengalese finch song lacks a single stereotyped sequence, and instead follows a more stochastic pattern. Thus, we separately analyzed global and local factors, and then assessed local variation after eliminating global changes.

Methods

Recording

The songs of eighteen male Bengalese finches (*Lonchura striata* var. *domestica*) were recorded. All birds were adult (>140 dph), and thus their songs had already crystallized. None of the birds shared tutor songs (i.e., individual songs were sufficiently different in note morphology and sequence pattern). Birdsongs were recorded by a microphone (Audio-technica, PRO35) that attached above a birdcage, which was placed in a soundproof chamber. Output from the microphone was amplified by a mixer (Mackie 402-VLZ3) and digitized via an audio interface (Roland UA-1010/UA-55) sampled at 44.1 kHz/16-bit. Birds were placed in the recording chamber at least 1 day before the start of recording, and were removed from the chamber the day after recording to obtain songs from an entire day. To reduce storage size, custom software was used to record only singing by monitoring the sound amplitude level. The recorded data were then down-sampled to 32 kHz. Birds were obtained from an aviary located in our laboratory at the University of Tokyo, Japan. The temperature and humidity of the aviary were maintained at approximately 25 °C and 60 %, respectively. Light/dark cycle was 13/11 h.

Note segmentation

To separate song notes from the successive recoding data, we used a thresholding algorithm as follows. First, an original waveform was bandpassed at 1–8 kHz and its amplitude envelope was extracted by full-wave rectification and lowpass filtered at 200 Hz. Then, periods of sound were detected by thresholding the amplitude envelope at a predefined level. We defined the level threshold as 6–10 SD above the mean of the background noise level. The mean of the background noise was estimated by detecting a peak in the sound level histogram. The SD was estimated from the full width at half maximum value of the histogram. If

two sounding elements had a micro-gap (duration typically less than 5 ms) between them, they were integrated into one note. The lower limit of gap duration was adjusted between 3 and 13 ms for each bird as necessary to prevent fragmentation of the note. Last, short elements whose durations were less than 10–30 ms (threshold adjusted for each bird) were regarded as simply motion noises or beak clicking sounds, and were not labeled as song notes. Suitable amplitude and duration thresholds were chosen for each bird by visual inspection of the spectrograms.

Note and gap labeling

After note segmentation, note types were classified by a semi-automatic procedure described in our previous article (Tachibana et al. 2014). First, we made an instruction data set that was used to train the machine classifier. The instruction data were derived from 2 min (~700 notes) of sound excerpts that were randomly extracted from various time points within the recording of each bird. All notes in the instruction dataset were manually labeled. The average number of note labels was 8.3 ± 2.59 (M ± SD) per bird. Non-singing calls, introductory notes, motion-related noises, and other environmental noises were all classified as elements of no interest (i.e., labeled as 'x'). Then, we calculated high-dimensional acoustic features (such as duration, entropy, zero cross, and spectral and cepstral coefficients) for all recorded notes. We trained a linear SVM classifier by the instruction dataset, and finally, the trained classifier classified all remaining data.

A gap was defined as a silent period with duration of less than 300 ms during singing. Gaps were labeled by a combination of preceding and following note types, e.g., the gap between notes 'a' and 'b' was labeled as 'a-b' (Fig. 1a). To eliminate possible erroneous labeling caused by the note classification error, we only labeled gaps existing between note pairs whose occurrence and transition probabilities were more than 1 %. All silent periods which were not labeled as a gap were labeled as elements of no interest (such as 'x-x'). The average number of gap labels was 9.3 ± 2.99 per bird.

A song typically begins with introductory notes and is followed by a resting period of several seconds. We termed the unit of each song rendition a 'bout'. Silent periods that were longer than 300 ms were regarded as boundaries of song bouts. The bouts consisting of less than two notes (irrespective of 'x' elements) were excluded from all analyses to avoid taking incomplete bouts into account.

Fine timing calculation

Because level thresholding did not provide sufficient temporal resolution for determining segment boundaries, we
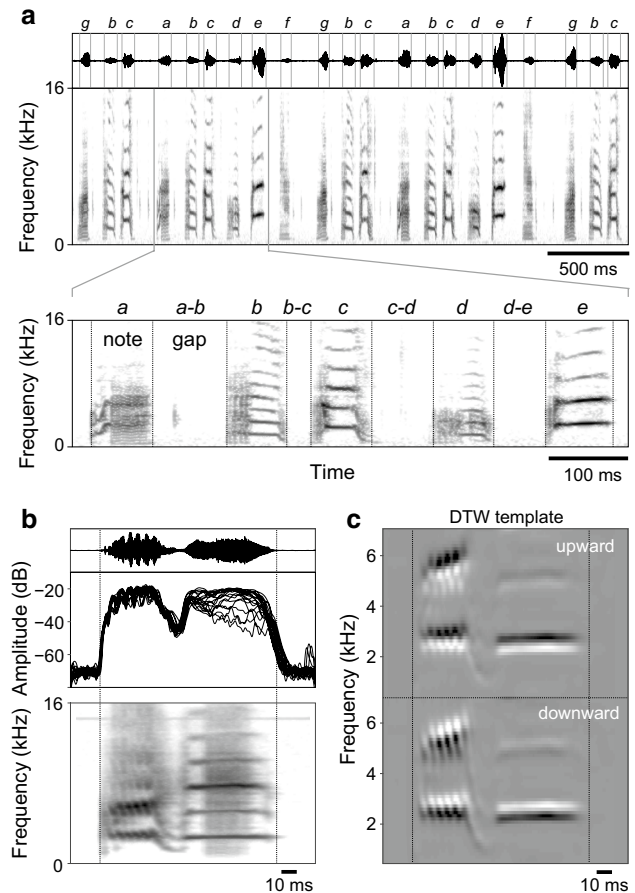

Fig. 1 Examples of spectrograms, waveforms and derivatives that show labels and note–gap boundaries. **a** Spectrogram of example song excerpt with notes and gaps labeled. **b** Amplitude envelopes and averaged spectrogram of one representative type of note. **c** Two derivatives of spectrogram used for dynamic time warping (DTW) to detect fine-grained temporal boundaries between notes and gaps. These were derived as diagonally upward and downward versions of spectro-temporal derivatives of original spectrogram shown in **b**. *Dotted lines* show boundaries of note onset and offset

employed dynamic time warping (DTW; see Anderson et al. 1996; Glaze and Troyer 2006, 2007) to determine the precise temporal boundaries between notes and gaps. The time warping was performed on spectro-temporal derivatives of the spectrogram. Original spectrograms of all same-type notes had 125 Hz and 0.5 ms resolutions (256-point Hanning window, 248-point overlap), and were normalized to have the same amplitude levels before calculating their derivatives. The spectrogram was convolved with a filter coefficient set [−2, −1, 0, 1, 2] in both vertical (i.e., frequency) and horizontal (time) directions to obtain 5-point gradient slopes for making the spectro-temporal derivatives. We made two types of the derivative; one was diagonally upward and the other was diagonally downward version, those were calculated by flipping the filter coefficient upside down. The vertical concatenation of upward and

downward derivatives (Fig. 1c) was used as a feature space in the DTW processing. Only spectral regions between 500 and 7000 Hz of the derivative were used. The warping template was defined as an average of the spectrogram derivatives of notes whose duration were between 25 and 75 % of all of the same-type notes. Onset and offset boundaries of the template were defined by visual inspection of derivative and original spectrograms. For each pair of time frames, the cosine similarity was calculated by measuring the distance between the template and each note. The template was time-warped and fitted to each note sample by a standard DTW algorithm. Then, the timing boundaries of the note were determined from the warped template. Outliers (determined as data more than 1.5 interquartile ranges below the first quartile or above the third quartile for each note and gap type) were excluded from further analyses. This outlier process excluded on average 2.6 ± 0.7 and 2.4 ± 0.7 % of notes and gaps, respectively, per bird. Finally, we collected 149 note and 168 gap types from all of the birds, and 1969.7 ± 1701.4 and 1313.4 ± 911.7 samples per note and gap type, respectively.

Overall properties

First, we assessed the overall distributions of mean and coefficient of variation (CV) of duration for each note and gap type. In these analyses, we assessed differences in the note and gap distributions using the Wilcoxon rank-sum test with significance level $\alpha = 0.05$. Moreover, we searched for differences in the temporal properties of gaps between stereotypic and branching transitions. Gaps were divided into two groups according to the transition probability of the preceding and following note pair surrounding the gap: the 'stereotypic' group was defined as gaps with a transition probability of $\geq$90 % (indicated as arrowheads and dotted lines in Fig. 3a, f), and the remaining gaps were placed in the 'branching' group. The transition probability was calculated in two ways, as both the divergence and convergence probabilities (Wohlgemuth et al. 2010; Fujimoto et al. 2011; Bouchard and Brainard 2013), where the divergence probability is a transition probability from one note to the following note, and the convergence probability is from one note to the previous note. We used a Wilcoxon rank-sum test with significance level $\alpha = 0.05$ to assess differences in the distribution of stereotypic and branching groups.

Within-day and within-bout variation analysis

To identify slowly varying global factors of song temporal control, we assessed within-day and within-bout variations before directly accessing local temporal features. For the within-day analysis, note and gap durations were converted

into proportional values (in percentages) for each type by dividing by the mean duration for that type. The inter-onset interval (IOI) was also calculated as a summed duration of the note and gap for each note pair, and converted into the proportional values. Then, data were averaged for each 1 h time bin. Time bins that contained less than 20 samples were discarded to ensure that average data were stable. The linear rate of change was calculated as slopes by linear regression for within-day changes of note, gap, and IOI durations. Statistical differences in the slopes was assessed by a one-sample t test with significance level $\alpha = 0.05$ (with Bonferroni correction for three comparisons). Additionally, hourly changes of sound amplitude (defined as the root-mean-squares (RMS) of note waveforms), bout duration, and note production rate were assessed. The change in amplitude and bout duration was calculated as the deviation from averaged values of all data for each bird. For the calculation of the note production rate, the number of notes was counted every hour and converted into a proportion of the total count.

We analyzed within-bout changes in a similar way to the within-day analysis, but within each bout. The proportional values (in percentage) of note, gap, and IOI durations were zero-centered by subtracting the mean values within each bout to eliminate between-bout differences (or hourly drift), and averaged for each one-second time bin. Time bins containing less than 20 samples were discarded to ensure that average data were stable. Then, the rate of change was calculated as slopes by linear regression for the note, gap, and IOI durations of each bird. For this calculation, we actually divided the data into two parts: an 'early' portion, including data up to 6.5 s after the bout started, and a 'late' portion which was the remaining data. We employed this method because results appeared to show different tendencies between the early and late portions of the bout (see Results section and Fig. 5). Only bouts whose durations were longer than 10 sec were included in the slope analysis to avoid possible biases in shorter bouts. Using this criterion, data from three birds were excluded from this analysis. The slopes of the linear regression obtained from fifteen birds were statistically assessed by a one-sample t test with significance level $\alpha = 0.05$ (with Bonferroni correction for six comparisons).

Cancelation of global factors and auto-regression analysis

The variation factors affecting local temporal structure were investigated after eliminating global factors, which resulted in slowly varying drift-like changes, such as song tempo acceleration/deceleration over song bouts. We performed this cancelation for each song bout using a cancelation line (see Fig. 6a) consisting of two slopes (for the early

and late parts of the bout) and one constant term (mean of each bout). These slopes and constant term corresponded to within-bout and within-day drift, respectively. The slopes were identical to the ones already estimated in the within-bout analysis for note, gap, and IOI data in the early and late parts of the bout for each bird. The constant values were calculated as the mean values within each bout. The cancelation line had proportional values (relative to the global mean of the entire data set). All note, gap, and IOI durations were divided by corresponding cancelation lines for each bout, respectively. We assessed the distributions of CVs for each note and gap type after the cancelation. The cancelation was performed on data for fifteen birds whose slopes were obtained as explained in the previous section, and we obtained 123 note and 138 gap types after the cancelation. We used a Wilcoxon rank-sum test with significance level $\alpha = 0.05$ to assess differences in the note and gap distributions.

Next, we estimated the auto-regression (AR) effects on song temporal structure. One can hypothesize a trade-off relationship among successive durations of neighboring elements to maintain the overall tempo of the bout. For example, a note that is prolonged from its average duration may result in the following note being shorter than its average duration, and vice versa. Furthermore, it can also be expected that this type of trade-off relationship could create small duration changes in nearby elements (i.e., successive notes) as well as elements that are more distant. To clarify if there exists such a self-regressive relationship between successive song elements, we performed a five-ordered AR analysis on the within-bout data. This analysis was separately performed on note, gap, and IOI duration sequences. Note that we used the data set in which global effects had been canceled. The mean value was subtracted from the original duration for each element to determine the deviation of duration. This residual duration stream underwent the AR analysis, which is expressed as follows:

$$y_n = x_n - \bar{x}_n$$

$$y_n = -a_1 y_{n-1} - a_2 y_{n-2} - \cdots - a_5 y_{n-5}$$

where x_n indicates the duration of the nth element (note, gap, or IOI) in a given song bout, and $\overline{x_n}$ is the mean value of element x_n across all song bouts. Thus, y_n represents the duration deviation of x_n, and a_1–a_5 are the AR coefficients. If the mth AR coefficient (a_m) has a positive large value, two elements that apart in distance of m have a negative correlation (trade-off). Additionally, we performed the AR analysis on a data stream that consisted of both notes and gaps (i.e., note–gap–note–gap…). For this note–gap stream, we estimated ten-ordered AR coefficients. Notes and gaps that could not be classified as a given type were replaced with random values with standard deviation that

was matched to the analyzing data stream. The successive introductory notes (repeated >3 times) were eliminated from this analysis because of their instability. We also eliminated short streams that included less than twenty notes to ensure reliable results. This limitation excluded data from one bird, and hence, we performed the analysis only on the remaining 14 birds. To remove artifacts in the AR coefficients that are usually caused by the limitation of shorter stream length, we made permuted streams where the duration deviations were same as the original stream, but randomly ordered, and subtracted AR coefficients of the permuted data from that of the original (not permuted) stream. Statistical differences of the relative AR coefficients were assessed by one-sample t tests with significance level $\alpha = 0.05$ (with Bonferroni correction). Note that trade-off effects between adjacent elements (first AR coefficient) in IOI or note–gap streams are possibly caused by measurement errors in detection of the boundary of note onsets and offsets. Thus, we did not focus on the first coefficient of IOIs and note–gap steams in the discussion section.

Results

Overall temporal properties and temporal properties related to transition probability

A total 149 note types and 168 gap types were derived from the 18 birds. On average, gaps (M ± SD: 45.8 ± 30.1 ms) were significantly shorter than notes (89.3 ± 28.9 ms); Wilcoxon rank-sum test, $z = 11.46$, $p < 0.001$ (Fig. 2a). A tendency in which longer duration was associated with larger variability was observed in both gaps and notes, though this tendency was greater in gaps than in notes (Fig. 2b). Mean distributions of note and gap durations, clearly demonstrated that gaps were more variable in duration than notes (Fig. 2c), even if having almost the same range of mean duration (60–80 ms, indicated as a black bar above Fig. 2b; 32 note and 30 gap labels). The coefficient of variation (CV) of most notes (3.7 ± 1.6 %) was significantly less than that of gaps (10.2 ± 4.1 %), Wilcoxon rank-sum test, $z = -14.38$, $p < 0.001$ (Fig. 2d). This difference in variability was similarly confirmed in the non-parametric version of variance estimates (i.e., quartile coefficients of dispersion): note (2.5 ± 1.1 %), gap (6.9 ± 3.0 %); Wilcoxon rank-sum test, $z = -14.34$, $p < 0.001$. Additionally, we calculated CVs of notes and gaps in a data set which was processed with a more relaxed outlier threshold of 3.0-interquartile range to test the influence of the outlier exclusion process. We derived a similar pattern of results: note (4.1 ± 1.8 %), gap (11.5 ± 4.4 %); Wilcoxon rank-sum test, $z = -14.51$, $p < 0.001$.
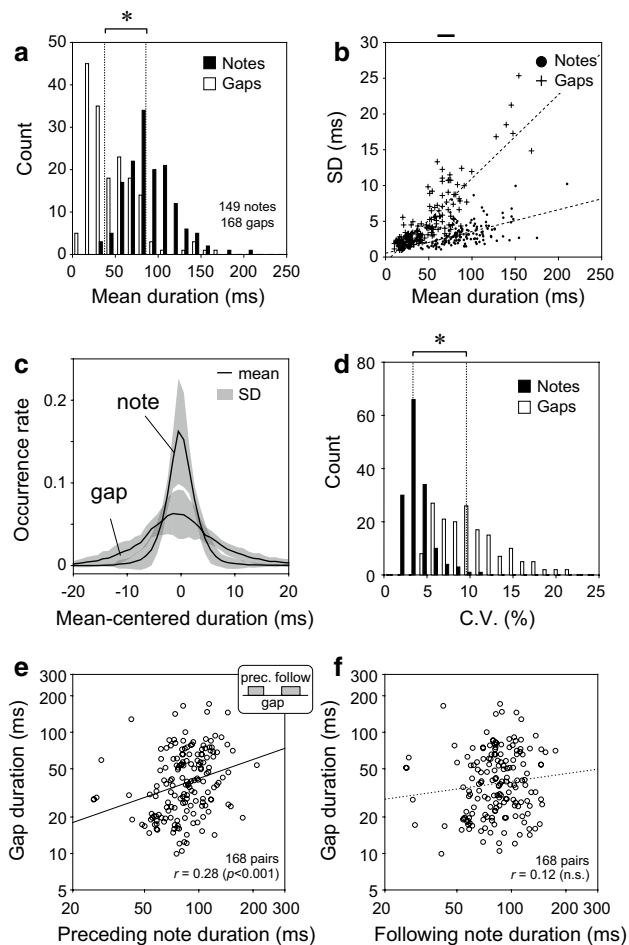
Fig. 2 Summarized profiles of mean and variability of note and gap durations. **a** Distributions of mean durations of respective classes (labels) of notes and gaps. *Dotted lines* show the median of the distributions. **b** Relationship between mean duration and standard deviation for notes and gaps. *Broken lines* represent fitted regression lines. **c** Average of note and gap duration distributions, which have nearly the same range of mean duration (60–80 ms), indicated as a *black bar* above **b**. *Gray hatched zone* shows the standard deviation. **d** Distributions of duration CVs (coefficients of variation) of respective note and gap types. *Dotted lines* show the median of distributions. In **a**, **d**, *asterisks* indicate significant differences, $p < 0.05$ (Wilcoxon rank-sum test, $n = 149$ notes, 168 gaps). Relationship between mean duration of a gap and that of preceding (**e**) and following (**f**) notes of that gap. *Bold and broken lines* are fitted regression lines

The relationship between a gap and its preceding or following note was also assessed. Each gap type was paired with a note type that preceded or followed that gap type. Then, mean durations of mean durations of gap types were plotted as functions of those note types on the logarithmic scale (Fig. 2e, f). The preceding notes had mean durations significantly correlated with mean gap duration: Pearson's $r = 0.28$, $n = 168$, $p < 0.001$; though the following notes did not: Pearson's $r = 0.12$, $n = 168$, $p = 0.137$.

Transition probability and gap duration

Differences in temporal properties between 'stereotypic' and 'branching' gaps were investigated. For divergence probability, stereotypic and branching groups had 60 and 108 gaps, respectively. Likewise, when measuring convergence probability, stereotypic and branching groups had 64 and 104 gaps, respectively. The mean gap duration of the stereotypic transition (divergence: 37.9 ± 24.6 ms; convergence: 37.9 ± 26.7 ms) was significantly shorter than that of the branching transition (div.: 50.1 ± 32.0 ms; con.: 50.6 ± 31.2 ms) for both divergence and convergence probabilities; Wilcoxon rank-sum test, div.: $z = -2.74$, $p = 0.012$ (Fig. 3c); con.: $z = -3.09$, $p = 0.004$ (Fig. 3h). However, the CV of gaps did not show significant differences between the stereotypic (div.: 9.6 ± 4.4 %; con.: 9.8 ± 4.5 %) and branching groups (div.: 10.5 ± 3.9 %; con.: 10.4 ± 3.8 %); Wilcoxon rank-sum test, div.: $z = -2.09$, $p = 0.073$ (Fig. 3e); con.: $z = -1.45$, $p = 0.295$ (Fig. 3j).

Within-day variations

Within-day trajectories of note durations and IOIs showed a similar pattern, having the peak around the first several hours after the light turned on, and decaying toward the evening (Fig. 4a, c). Mean duration per an hour of proportional note durations showed a tendency to shorten gradually from morning to evening while gap durations did not show a clear pattern (Fig. 4b). The slopes of linear regression fitting (Fig. 4d) for note durations and IOIs as functions of time were significantly negative (-0.14 ± 0.08 and -0.11 ± 0.10 percentages per hour, respectively) but this was not the case for gap durations (-0.03 ± 0.17): one-sample t test with Bonferroni correction; note: $t(17) = -7.34$, $p < 0.001$; gap: $t(17) = -0.77$, $p = 1.261$; IOI: $t(17) = -4.31$, $p = 0.001$. To assess a coupling between a note's duration and its amplitude we also measured the within-day change of sound amplitude in notes (Fig. 4e), however, no clear tendency was found across birds. Note production rates showed a variable, but a weak tendency towards declining from morning to evening (Fig. 4f). Bout durations did not show obvious change throughout the day (Fig. 4g).

Within-bout variations

Regarding variations within bouts, the note duration and IOI tended to be prolonged gradually (Fig. 5a, c), while the gap duration did not show a consistent change (Fig. 5b). More specifically, the gap duration appeared to be gradually shortened and the IOI was almost constant
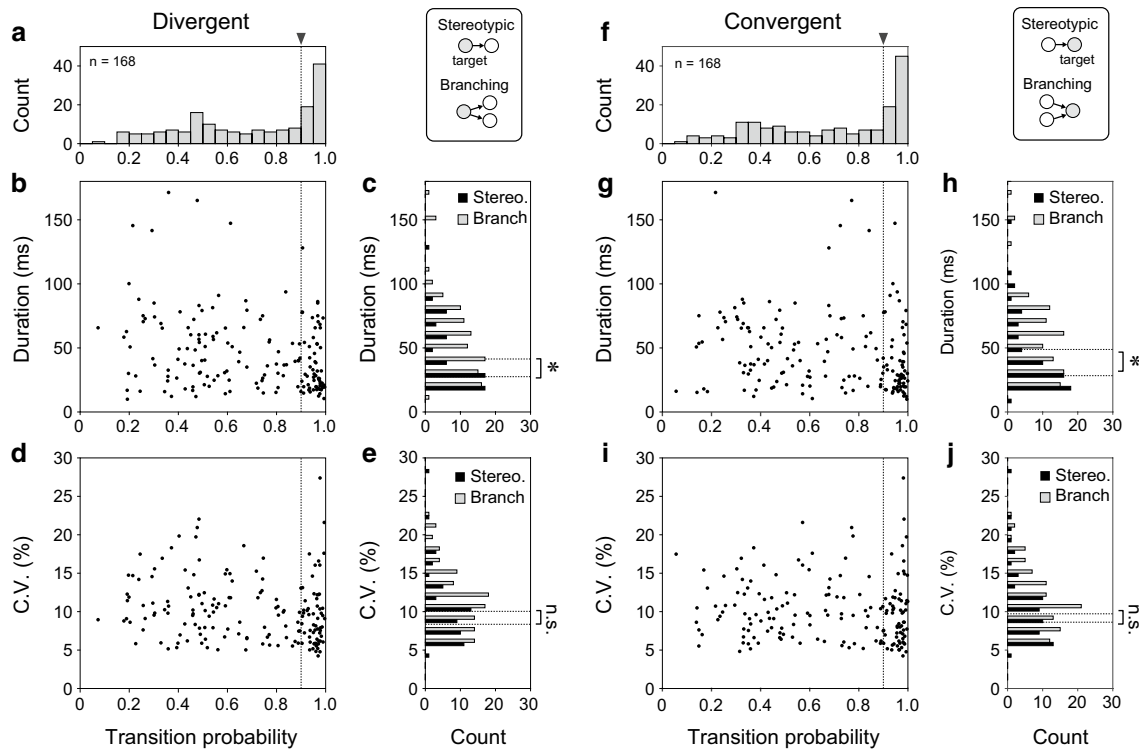
Fig. 3 Relationship between gap duration and two types of transition probability (divergent and convergent transition). **a** Histogram showing distribution of divergent transition probabilities of 168 types of gap. *Scatter and bar plots* showing relationship between transition probability and mean duration of gaps. In the *bar plot* (**c**), gaps were divided into two groups according their transition probabilities: stereotypic (*black bars*) or branching (*gray*) transitions. The stereotypic transition was defined as a transition with the divergent probability of ≥90 %, indicated as an *arrowhead* and *dotted line* in **a**. The remaining transitions were classified as branching. **d, e** *Scatter and bar plots* showing relationship between transition probability and CVs of gaps, similar to **b, c**. *Dotted lines* in the bar plots show the median of distributions, and *asterisk* indicates significant differences, $p < 0.05$ (Wilcoxon rank-sum test, Bonferroni corrected). **f–j** are the same as **a–e**, but using the convergent version of transition probability

within the early part of bout (until 6.5 s). In the late part of bout (after 6.5 s), the gap duration and IOI tended to be prolonged. We separately performed the linear regression fitting to both the early and late parts (1–6.5 and 6.5–14 s) of the 14 s bout data to confirm the differences observed in gap and IOI durations for early and later portions of the song bout. We found that (Fig. 5d) the early data were significantly positive only for notes (0.31 ± 0.19 percentages per second), but not for gaps (-0.40 ± 0.77) or IOIs (0.07 ± 0.28); one-sample t test with Bonferroni correction; notes: $t(17) = 6.33$, $p < 0.001$; gaps: $t(17) = -2.03$, $p = 0.372$; IOIs: $t(17) = 0.99$, $p = 2.033$. However, the slopes for the late parts of song were significantly positive for notes (0.16 ± 0.09), gaps (0.42 ± 0.24), and IOIs (0.26 ± 0.12: notes: $t(17) = 6.42$, $p < 0.001$; gaps: $t(17) = 6.81$, $p < 0.001$; IOIs: $t(17) = 8.55$, $p < 0.001$. Additionally, the note amplitude (calculated as RMS of the note waveform) was increased during the bout (early: 0.38 ± 0.29 dB/s, late: 0.11 ± 0.14 dB/s; Fig. 5e). Mean bout duration was 9.86 ± 4.04 s (Fig. 5f), and the mean

number of notes in a bout was 65.75 ± 35.3 (figure not shown).

Local variation

To focus on local temporal features, the slow drifting components, i.e., within-day and within-bout variation, were canceled out from the song timing dataset (as described in the Method section). Cancelation of global and middle range factors slightly decreased the CV of notes (3.5 ± 1.6 %) and gaps (10.0 ± 3.9 %), but the relative difference between them remained the same (Fig. 6b; see also Fig. 2d). The CVs of gaps were significantly greater than that of notes, Wilcoxon rank-sum test, $z = -13.06$, $p < 0.001$.

On this pre-processed dataset, we assessed trade-off relationships among durations of local neighboring elements. We performed the AR analysis of duration deviation within bouts to quantify how the trade-off effect was localized in the timing sequence. The analysis was separately performed on note, gap, and IOI streams, and also on
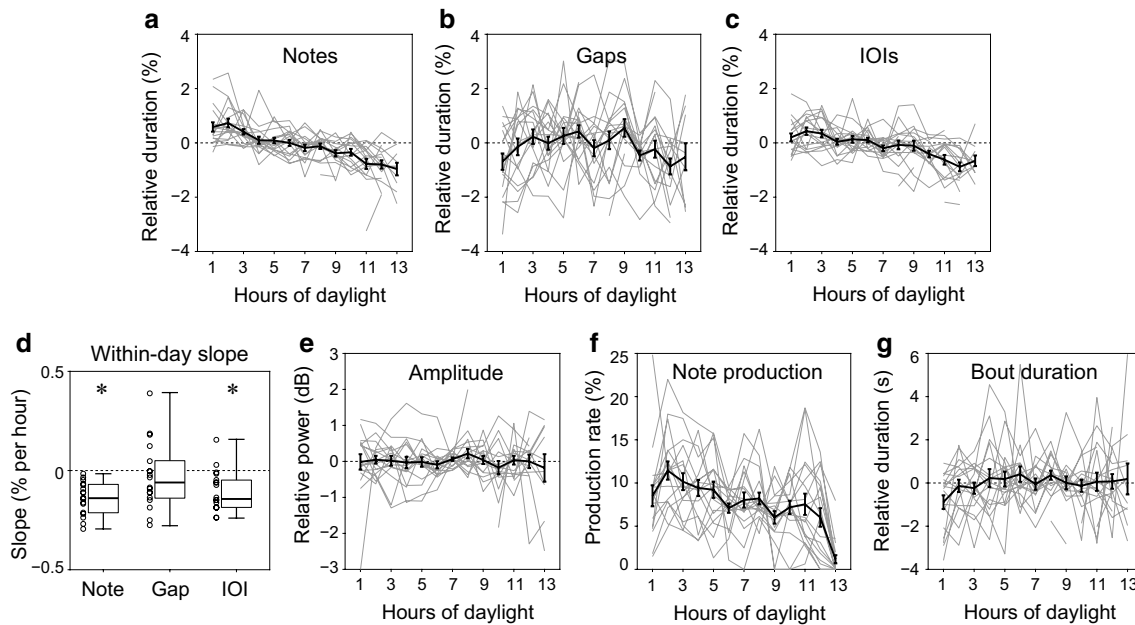
Fig. 4 Within-day variations. Normalized duration (%) of notes (**a**), gaps (**b**), and IOIs (**c**) derived using time bins of 1 h, during which recording chamber lights were on. **d** Linear slopes of within-day slow drift of note and gap durations. The slope was calculated by a linear fitting to the normalized duration data. *White circles* indicate the slope value for each bird. *Asterisks* indicate significant differences, $p < 0.05$ (one-sample t test, Bonferroni corrected). **e** Normalized power amplitude (dB) during daytime with same time bins and steps as in **a–c**. **f** Proportional rates (%) of note production during daytime. **g** Changes in bout duration (in seconds) during daytime. In **a–c**, **e–g**, thin *gray lines* represent data from each bird. *Bold black lines* indicate the means and standard errors ($n = 18$ birds)

successive note–gap sequences. Note that first coefficients obtained from AR analysis of IOI and note–gap streams were possibly due to measurement errors in detection of note boundary. Hence, they are not discussed in detail. The results showed no significant trade-off effects in note, gap and IOI stream (Fig. 6c–e) with the exception of the first coefficient for the note–gap stream (Fig. 6f). Statistical tests confirmed that only the first coefficient for note–gap stream was significantly positive; one-sample t test with Bonferroni correction, a_1(note–gap): $t(13) = 4.28$, $p = 0.009$; a_{2-10} (note–gap): $t(13) = -2.52 - +2.38$, $p = 0.33 - 9.631$.

Discussion

Using one-day-long recordings of Bengalese finch songs, the current study showed evidence in favor of distinct multiple sources of temporal variability in motor sequences. The duration of silent intervals, or gaps, was highly variable compared to that of song notes. Mean gap durations were correlated with that of preceding notes, and also with transition probability. Characteristic tendencies were found in three time ranges: within-day drift, within-bout changes, and local jitter. These results suggest that there are distinct sources of temporal variability at multiple levels on the basis of the note–gap duration relationship.

General properties and differences in notes and gaps

We found that gaps had 3.1 times more variability than notes (2.8-times after cancelation of slow drifts). A similar variability difference has been reported in note and gap durations of zebra finches (Glaze and Troyer 2006, 2012; Cooper and Goller 2006; Andalman et al. 2011), though the differences in our results appear to be slightly larger than those of previous reports. In zebra finch studies, it has been suggested that the variability difference between sound elements and gaps was caused by different control mechanisms (Andalman et al. 2011); it is likely that HVC controls expiratory activities during the production of sound elements, but for gap production, inspiratory activity could be influenced by noisy signals in circuits downstream of HVC.

Interestingly, the mean duration of a gap was correlated with the mean duration of the note preceding the gap, but not with that of the note following the gap. This note–gap relationship suggests that the timing control of gaps and their preceding notes is associated with one other in song production system. One zebra finch study suggested that inter-onset interval (preceding note + gap), but not inter-offset (gap + following note), represents the preferred alignment unit in song tempo control (Glaze and Troyer 2006) according to their results on trade-off relationships
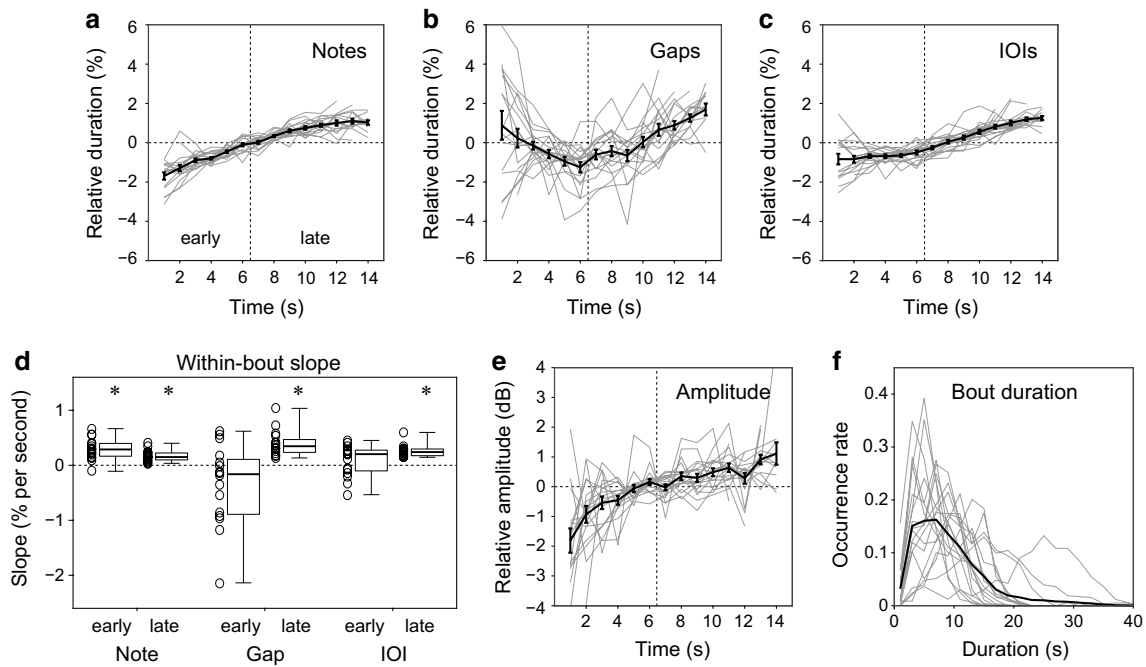
Fig. 5 Within-bout variations. Normalized duration (%) of notes (**a**), gaps (**b**), and IOIs (**c**) derived using time bins of one second, in the bout. *Thin gray lines* represent data from each bird. *Bold black lines* indicate the means and standard errors ($n = 18$ birds). The *vertical broken line* indicates the boundary between early (1–6 s) and late (7–14 s) periods of song, which appeared to show different temporal patterns. **d** Linear slopes of within-bout changes in note and gap durations. The slopes were calculated from data in the early and late periods separately. *Open circles* indicate the slope value for each bird. *Asterisks* indicate significant differences, $p < 0.05$ ($n = 15$ birds; one-sample t test; Bonferroni corrected). **e** Normalized power amplitude (dB) during the bout, plotted with the same manner as **a**–**c**. **f** Overall distribution of bout durations

between note and gap durations. On the other hand, a study on song copying in zebra finches reported that when finches copy two chunks of song from a tutor, the durations of between-chunk gaps of the learned song appear to be more similar to the gap that precedes the second chunk rather than the gap that follows the first chunk (Williams and Staples 1992). This suggests the presence of inter-offset, but not inter-onset, units in the song development system. The inconsistency between this study and our results might be associated with the dissociation in song control units between the perception of tutor songs for learning and the production of the bird's own song. To clarify this issue, it would be helpful to perform a variability analysis on a combined data set made of the tutor songs and the developed songs of the juveniles.

Transition probability and gap duration

The observed difference in gap duration between the stereotyped and branching transitions could be interesting in considering the hierarchical structures of song motor control. This tendency was common in the two types of data derived using the divergence and convergence probabilities, although one study suggests that the convergence transition is more

naturalistic in neural circuits than the divergence transition (Bouchard and Brainard 2013). The gap durations in stereotypic transitions were relatively short, whereas branching gaps had a variety of durations. This could be related to the finding in zebra finches that 'motif-boundary' gaps were relatively longer and more variable than 'within-motif' gaps (Glaze and Troyer 2006, 2012), though our data did not show significant differences in the variability. In the Bengalese finch songs, we could not define a 'motif' or stereotypic subsequences, due to of the variable, stochastic nature of song sequencing (Okanoya 2004). However, several previous studies in the Bengalese finch suggest the existence of note 'chunks' which are stereotypic transitions of several notes in song production (Seki et al. 2008; Takahasi et al. 2010) and in song perception (Suge and Okanoya 2010). It has also been found that 'within-chunk' gaps were shorter than 'between-chunk' gaps (Takahasi et al. 2010). Taken together, these studies and our results consistently suggest that the transition stereotypy or chunking structure has some association with gap duration.

Acceleration throughout the day

The gradual changes during a day in note durations and IOIs are consistent with previous reports that demonstrate
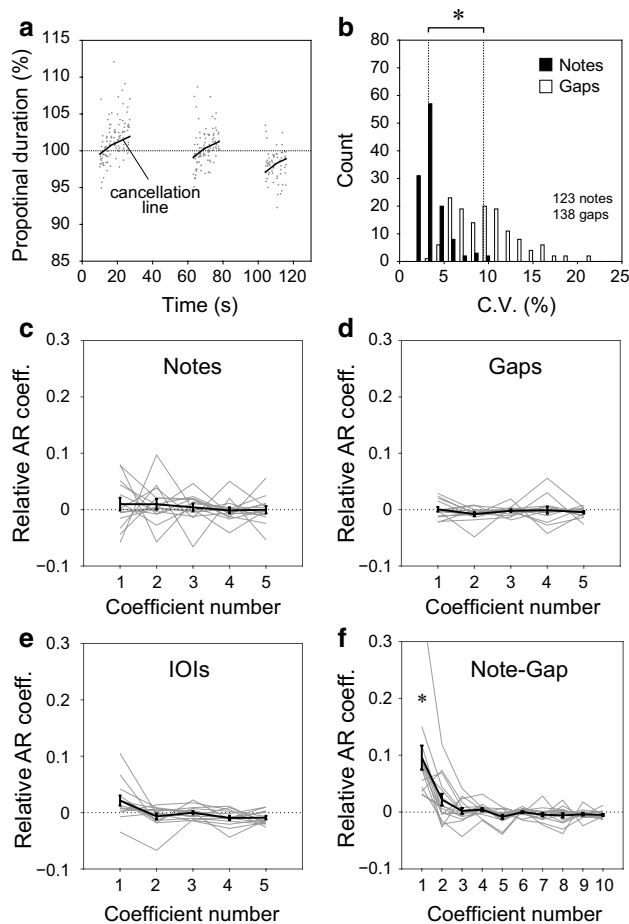
Fig. 6 Cancelation of global drift and local trade-off effects. **a** Example of cancel lines for global drift cancelation. *Gray dots* show proportional duration of notes and *black lines* indicate the cancel lines. **b** Distributions of duration CVs (coefficients of variation) of respective note and gap classes after the cancelation. *Dotted lines* show the median of the distributions, and *asterisks* indicate significant differences, $p < 0.05$ (Wilcoxon rank-sum test). **c–e** Coefficients of the auto-regression (AR) analysis on the duration deviation of notes (**c**), gaps (**d**), IOIs (**e**), and note–gap stream (**f**) within the bout. *Thin gray lines* represent data from each bird. *Bold black lines* indicate the means and standard errors ($n = 14$ birds). The AR coefficients were shown as deviations from corresponding permuted data

a circadian change in song tempo (Glaze and Troyer 2006, 2012; Aronov and Fee 2012). However, the observation that the within-day pattern in note durations and IOIs of the Bengalese finch song was inconsistent with the pattern of change in song tempo of zebra finches, which speeds up during the initial several hours of the day, and then slows down until evening (Glaze and Troyer 2006, 2012). This discrepancy in the pattern of change between them could reflect species differences in the timing control of the song motor system. On the other hand, another explanation can be raised from differences in body temperatures based on findings that demonstrate that the song tempo depends on

the temperature in the HVC nucleus (Long and Fee 2008; Andalman et al. 2011). In fact, it has been suggested that the circadian change of the song tempo could be explained by the daily pattern of the brain temperature in zebra finches (Aronov and Fee 2012). Thus, the observed within-day change in note durations and IOIs could reflect the circadian changes in body temperature.

Slowing during a single bout

We found an effect such that the note duration clearly became prolonged within each bout. This within-bout increase in duration was accompanied by an increase in note amplitude. This coupling of increments in note duration and amplitude observed within the bout was not observed in the day-long data, as mentioned above. This suggests different mechanisms underlying the observed duration drift in these two different time ranges. A previous study in zebra finches (Chi and Margoliash 2001) recorded neurons in the premotor nucleus (RA) during singing (11 neurons from 4 birds). The authors found a similar note prolongation as we found, associated with the firing pattern in RA. This suggests that the observed prolongation could be attributed to changes in premotor neural activities. The consistency of our findings demonstrates that the prolongation is general to a larger sample over different species. The covariation of note duration and amplitude has been observed in the introductory notes of zebra finches, and could be interpreted as an index of excitation or "readiness" state for singing (Rajan and Doupe 2013). Although, contrary to our data, note duration in zebra finches shrinks with the amplitude increment. However, our data did not include the introductory notes, and observed changes were rather mild compared to the previous report. The gap duration showed decrement complementarily to the note prolongation during the first 6.5 s in the bout, and this was reflected in the relatively constant IOI. This complementary changes between notes and gaps during the first half part of the bout could be explained by increasing internal excitation, or arousal, under a constant tempo which causes a gradual increment of expiratory activity, and hence, increases both the note duration and amplitude. On the other hand, all of the notes, gaps and IOIs were prolonged during the latter part of the bout. One could assume that this tempo slowing might reflect the brain temperature changes (Aronov and Fee 2012) during a song bout, similar to the above-mentioned within-day changes. Or, the within-bout drifts in temporal properties might be attributed to accumulation of fatigue in the muscular and/or neuronal systems for singing. The potential mechanisms that cause the slowing down of song tempo during the song bout should be tested in future studies.

The sources of these two observed effects, within-bout prolongation and within-day shrinkage of note durations, are likely independent. The results of within-bout analysis suggest that mean note duration becomes shorter when bout duration is shorter. However, actual bout durations were almost constant throughout the day or prolonged slightly in the first several hours, whereas the note duration gradually shortened throughout the day. Thus, this discrepancy suggests independent sources for the drifts in these two different time ranges.

Local trade-off relationships

Our results showed a local trade-off effect between notes and gaps after removing slowly varying drift-like changes, such as song tempo acceleration/deceleration over and within song bouts. We used an auto-regression model to analyze how duration deviations of preceding elements had an influence on the current element. The model suggested that no obvious effect was found among adjacent elements in note, gap, and IOI dataset, with the exception of the last element (the one preceding the current element), which had a negative correlation in the note–gap sequence. Additionally, weak effects (but not significant) are associated with the first coefficients in IOI stream, and the second coefficients in note–gap stream (Fig. 6e, f). The first coefficients of auto-regression in IOI and note–gap dataset are possibly attributed to measurement errors in note boundary detection, and so we could not discuss this result as an intrinsic effect caused by bird's song production system. The weak increase in the second coefficient in the note–gap data is rather interesting, and might suggest local trade-off relationships between neighboring notes or gaps. However, the direct analysis on note or gap stream did not show such an increase in the first coefficient.

At the very least, these results would suggest that local variability has no obvious relationships among neighboring elements after elimination of global effects (within-day and within-day drift), and this local jitter might be caused by a random source generated in song motor systems. These temporal variabilities at multiple levels constitute temporal structures on the basis of the note–gap duration relationship in Bengalese finch song.

References

Amador A, Perl YS, Mindlin GB, Margoliash D (2013) Elemental gesture dynamics are encoded by song premotor cortical neurons. Nature 495:59–64. doi:10.1038/nature11967

Andalman AS, Fee MS (2009) A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. Proc Natl Acad Sci USA 106:12518–12523. doi:10.1073/pnas.0903214106

Andalman AS, Foerster JN, Fee MS (2011) Control of vocal and respiratory patterns in birdsong: dissection of forebrain and brainstem mechanisms using temperature. PLoS One 6:e25461. doi:10.1371/journal.pone.0025461

Anderson SE, Dave AS, Margoliash D (1996) Template-based automatic recognition of birdsong syllables from continuous recordings. J Acoust Soc Am 100:1209. doi:10.1121/1.415968

Aronov D, Fee MS (2012) Natural changes in brain temperature underlie variations in song tempo during a mating behavior. PLoS One 7:e47856. doi:10.1371/journal.pone.0047856

Bouchard KE, Brainard MS (2013) Neural encoding and integration of learned probabilistic sequences in avian sensory-motor circuitry. J Neurosci 33:17710–17723. doi:10.1523/JNEUROSCI.2181-13.2013

Chi Z, Margoliash D (2001) Temporal precision and temporal drift in brain and behavior of zebra finch song. Neuron 32:899–910. doi:10.1016/S0896-6273(01)00524-4

Cooper BG, Goller F (2006) Physiological insights into the social-context-dependent changes in the rhythm of the song motor program. J Neurophysiol 95:3798–3809. doi:10.1152/jn.01123.2005

Fee MS, Goldberg JH (2011) A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. Neuroscience 198:152–170. doi:10.1016/j.neuroscience.2011.09.069

Fee MS, Scharff C (2010) The songbird as a model for the generation and learning of complex sequential behaviors. ILAR J 51:362–377. doi:10.1093/ilar.51.4.362

Fujimoto H, Hasegawa T, Watanabe D (2011) Neural coding of syntactic structure in learned vocalizations in the songbird. J Neurosci 31:10023–10033. doi:10.1523/JNEUROSCI.1606-11.2011

Glaze CM, Troyer TW (2006) Temporal structure in zebra finch song: implications for motor coding. J Neurosci 26:991–1005. doi:10.1523/JNEUROSCI.3387-05.2006

Glaze CM, Troyer TW (2007) Behavioral measurements of a temporally precise motor code for birdsong. J Neurosci 27:7631–7639. doi:10.1523/JNEUROSCI.1065-07.2007

Glaze CM, Troyer TW (2012) A generative model for measuring latent timing structure in motor sequences. PLoS One 7:e37616. doi:10.1371/journal.pone.0037616

Hahnloser RHR, Kozhevnikov A, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature 419:65–70. doi:10.1038/nature00974

Kao MH, Doupe AJ, Brainard MS (2005) Contributions of an avian basal ganglia–forebrain circuit to real-time modulation of song. Nature 433:638–643. doi:10.1038/nature03127

Leonardo A, Fee MS (2005) Ensemble coding of vocal control in birdsong. J Neurosci 25:652–661. doi:10.1523/JNEUROSCI.3036-04.2005

Long MA, Fee MS (2008) Using temperature to analyse temporal dynamics in the songbird motor pathway. Nature 456:189–194. doi:10.1038/nature07448

Long MA, Jin DZ, Fee MS (2010) Support for a synaptic chain model of neuronal sequence generation. Nature 468:394–399. doi:10.1038/nature09514

Okanoya K (2004) The Bengalese finch: a window on the behavioral neurobiology of birdsong syntax. Ann N Y Acad Sci 1016:724–735. doi:10.1196/annals.1298.026

Rajan R, Doupe AJ (2013) Behavioral and neural signatures of readiness to initiate a learned motor sequence. Curr Biol 23:87–93. doi:10.1016/j.cub.2012.11.040

Sakata JT, Brainard MS (2006) Real-time contributions of auditory feedback to avian vocal motor control. J Neurosci 26:9619–9628. doi:10.1523/JNEUROSCI.2027-06.2006

Scharff C, Nottebohm F (1991) A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. J Neurosci 11:2896–2913

Seki Y, Suzuki K, Takahasi M, Okanoya K (2008) Song motor control organizes acoustic patterns on two levels in Bengalese finches (*Lonchura striata* var. *domestica*). J Comp Physiol A Neuroethol Sens Neural Behav Physiol 194:533–543. doi:10.1007/s00359-008-0328-0

Sober SJ, Brainard MS (2012) Vocal learning is constrained by the statistics of sensorimotor experience. Proc Natl Acad Sci USA 109:21099–21103. doi:10.1073/pnas.1213622109

Sober SJ, Wohlgemuth MJ, Brainard MS (2008) Central contributions to acoustic variation in birdsong. J Neurosci 28:10370–10379. doi:10.1523/JNEUROSCI.2448-08.2008

Suge R, Okanoya K (2010) Perceptual chunking in the self-produced songs of Bengalese finches (*Lonchura striata* var. *domestica*). Anim Cogn 13:515–523. doi:10.1007/s10071-009-0302-4

Suthers RA, Margoliash D (2002) Motor control of birdsong. Curr Opin Neurobiol 12:684–690. doi:10.1016/S0959-4388(02)00386-0

Tachibana RO, Oosugi N, Okanoya K (2014) Semi-automatic classification of birdsong elements using a linear support vector machine. PLoS One 9:e92584. doi:10.1371/journal.pone.0092584

Takahasi M, Yamada H, Okanoya K (2010) Statistical and prosodic cues for song segmentation learning by Bengalese finches (*Lonchura striata* var. domestica). Ethology 116:481–489. doi:10.1111/j.1439-0310.2010.01772.x

Tumer EC, Brainard MS (2007) Performance variability enables adaptive plasticity of "crystallized" adult birdsong. Nature 450:1240–1244. doi:10.1038/nature06390

Warren TL, Charlesworth JD, Tumer EC, Brainard MS (2012) Variable sequencing is actively maintained in a well learned motor skill. J Neurosci 32:15414–15425. doi:10.1523/JNEUROSCI.1254-12.2012

Williams H, Staples K (1992) Syllable chunking in zebra finch (Taeniopygia guttata) song. J Comp Psychol 106:278–286. doi:10.1037/0735-7036.106.3.278

Wohlgemuth MJ, Sober SJ, Brainard MS (2010) Linked control of syllable sequence and phonology in birdsong. J Neurosci 30:12936–12949. doi:10.1523/JNEUROSCI.2690-10.2010

Woolley SC, Kao MH (2015) Variability in action: contributions of a songbird cortical–basal ganglia circuit to vocal motor learning and control. Neuroscience 296:39–47. doi:10.1016/j.neuroscience.2014.10.010

Yu AC, Margoliash D (1996) Temporal hierarchical control of singing in birds. Science 273:1871–1875. doi:10.1126/science.273.5283.1871