A statistical method for quantifying songbird phonology and syntax

Wei Wu [a,*], John A. Thompson [b,c], Richard Bertram [b,d], Frank Johnson [b,c]

[a] Department of Statistics, Florida State University, Tallahassee, FL 32306-4330, United States
[b] Program in Neuroscience, Florida State University, Tallahassee, FL 32306-1270, United States
[c] Department of Psychology, Florida State University, Tallahassee, FL 32306-4301, United States
[d] Department of Mathematics, Florida State University, Tallahassee, FL 32306-4510, United States

ARTICLE INFO

ABSTRACT

Songbirds are the preeminent animal model for understanding how the brain encodes and produces learned vocalizations. Here, we report a new statistical method, the Kullback–Leibler (K–L) distance, for analyzing vocal change over time. First, we use a computerized recording system to capture all song syllables produced by birds each day. Sound Analysis Pro software [Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, Mitra PP. A procedure for an automated measurement of song similarity. Anim Behav 2000;59:1167–76] is then used to measure the duration of each syllable as well as four spectral features: pitch, entropy, frequency modulation, and pitch goodness. Next, two-dimensional scatter plots of each day of singing are created where syllable duration is on the *x*-axis and each of the spectral features is represented separately on the *y*-axis. Each point in the scatter plots represents one syllable and we regard these plots as random samples from a probability distribution. We then apply the standard information-theoretic quantity K–L distance to measure dissimilarity in phonology across days of singing. A variant of this procedure can also be used to analyze differences in syllable syntax.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Zebra finches (*Taeniopygia guttata*) are a species of passerine songbird that learns and produces a motif, stereotyped sequences of 3–7 harmonically complex syllables. Analyses of both vocal development and experimental manipulation of adult motifs focus on vocal features that individuate birds; these include the sequential order and acoustic structure of syllables. Here, we present a method for quantitative, syllable-level analysis of motif syntax and phonology. This quantitative analysis combines a commonly used tool from information theory (Kullback–Leibler distance; Kullback and Leibler, 1951) with the Feature Batch module in Sound Analysis Pro (SA+; Tchernichovski et al., 2000) to quantitate comparison of motifs.

The Similarity Batch module of SA+ is a widely used standard for quantifying differences in birdsong (Hough and Volman, 2002; Wilbrecht et al., 2002a,b; Woolley and Rubel, 2002; Funabiki and Konishi, 2003; Liu et al., 2004; Zevin et al., 2004; Cardin et al., 2005; Coleman and Vu, 2005; Heinrich et al., 2005; Kittelberger and Mooney, 2005; Liu and Nottebohm, 2005; Olveczky et al., 2005; Phan et al., 2006; Teramitsu and White, 2006; Wilbrecht et al., 2006; Haesler et al., 2007; Hara et al., 2007; Liu and Nottebohm, 2007; Pytte et al., 2007; Roy and Mooney, 2007; Thompson and Johnson, 2007; Zann and Cash, 2008). This analysis is typically used to compare the acoustic properties of the motif across juvenile development, or before and after an experimental manipulation. Although Similarity Batch provides an efficient strategy to objectively analyze acoustic similarity across a large number of motifs, the inability to identify unique syllable contributions to the overall similarity score represents a general limitation for motif-based comparison. For example, if the comparison of two motifs results in a low similarity score, one does not know whether the low similarity resulted from a mismatch in the temporal order or phonological structure, or which syllables specifically contributed to the mismatch.

While other methods exist to quantify the acoustic properties of individual syllables (e.g., Tchernichovski et al., 2000, 2001; Deregnaucourt et al., 2005; Sakata and Brainard, 2006; Crandall et al., 2007; Ashmore et al., 2008), the primary advantage of our method is the ability to make multi-dimensional comparison of large numbers of identified syllables, generate an overall motif similarity value, and identify individual syllable contributions to motif syntax and phonology. In order to make syllable-level comparisons, we also present an empirical strategy to identify syllable clusters

* Corresponding author at: Department of Statistics, Florida State University, Tallahassee, FL 32306-4330, United States. Tel.: +1 850 644 9855; fax: +1 850 644 5271.
E-mail address: wwu@stat.fsu.edu (W. Wu).

(i.e., repeated instances of the same syllable). Other methods to identify syllable clusters are available – for example, the Clustering Module in SA+ which identifies individual syllables based on Euclidean distance in acoustic feature space, or the KlustaKwik algorithm used by Crandall et al. (2007). However, the method presented here can be applied no matter the syllable cluster strategy employed.

Our method begins by utilizing the Feature Batch module in SA+ to partition a large number of motifs (e.g., those produced by an individual bird during a day of singing) into syllables from which we generate 2D scatter plots of syllable acoustic features vs. the syllable duration. We then apply the K–L distance to quantify the degree of dissimilarity between populations of syllables derived from two different time points (or two different 2D scatter plots): one from a baseline day of singing and the second from any other sample or day of singing. Thus, the K–L distance is an effective tool for performing a multi-dimensional comparison of syllable acoustic features from one day of singing with song from any other day. We also describe how the K–L distance can be used to determine changes in the syntax of syllables within the motif. Both applications of the K–L distance would be useful approaches for comparing the motifs produced by a bird at different stages of development or prior to and after a manipulation such as ablation, infusion of pharmacological agents or altered sensory feedback.

2. Methods

2.1. Subjects and apparatus

Adult (>125 days) male zebra finches were individually housed in medium-sized bird cages (26.67 cm × 19.05 cm × 39.37 cm) placed within computer-controlled environmental chambers (75.69 cm × 40.39 cm × 47.24 cm). The environmental chambers prevented visual as well as auditory access to other birds. A computer maintained both the photoperiod (14/10 h light/dark cycle) and ambient temperature (set to 26 °C) within each chamber. Birds were provisioned daily with primarily millet-based assorted seed and tap water. Birds acclimated to the environmental chambers for 2 weeks before initiation of baseline recordings. All daily care procedures and experimental manipulations of the birds were reviewed and approved by the Florida State University Animal Care and Use Committee.

2.2. Surgery

HVC microlesion: birds were first deeply anesthetized with Equithesin (0.04 cc) and then secured in a stereotaxic instrument. The skull was exposed by centrally incising the scalp and retracting the folds with curved forceps. Following application of avian saline (0.75 g NaCl/100 mL dH$_2$O) to the exposed area, small craniotomies were placed over the approximate location of HVC bilaterally. To determine the locations of these nuclei, the bifurcation at the midsagittal sinus was used as stereotaxic zero. Bilateral HVC microlesions were performed by positioning an electrode (Teflon insulated tungsten wire electrode, with a 200 μm diameter; A-M Systems, Everett, WA, www.a-msystems.com) along the anterior–posterior axis directly lateral from stereotaxic zero with three penetrations per side beginning at 2.1 mm. The second and third lesions were at an interval distance of 0.4 mm, each with a depth of 0.6 mm. For each penetration, current was run at 100 μA for 30 s. Following the lesion, the incision was treated with an antiseptic, sealed with veterinary adhesive, and the bird returned to its home cage.

2.3. Data acquisition

Birds were maintained in complete social isolation for the duration of recording, thus only "undirected" songs (i.e., not directed toward a female) were recorded and analyzed. For all birds, song production was recorded in 24 h blocks using a unidirectional microphone fastened to the side of the internal cage. Sounds transmitted by the microphone were monitored through a computer that was running sound-event triggered software (Avisoft Recorder; Avisoft Bioacoustics, Berlin, Germany).

Song was captured in bouts (2–7 s bursts of continuous singing during which the motif may be repeated one to five times) and each song bout was saved as a time-stamped .wav file (44 kHz sampling rate) onto the computer hard drive where each day of singing by each bird was saved under a single directory. To insure the collection of all song bouts, we biased the triggering settings for Avisoft Recorder in favor of false-positive captures (i.e., .wav files composed of repeated calls and/or cage noises). Therefore, the file directories for each day of singing by each bird required selective deletion of false-positive .wav files.

Detection and removal of false-positive .wav files from the file directories involved a three-step process. First, we used Spectrogram (version 13.0; Visualization Software LLC) to convert all sound files (.wav) recorded during a single day of singing into frequency spectrograph image files (.jpg). Second, using an image-viewer and file management program (IMatch; M. Westphal, Usingen, Germany), we removed all .jpg images that contained singing-related spectra from each directory. We used the content-based image retrieval module in IMatch to streamline the detection and removal process. Briefly, users select a .jpg image containing singing-related spectra and this module reorders thumbnail images of all .jpg files in the directory by similarity; users then select images that contain singing-related spectra for removal. Visual inspection allows rapid and unambiguous discrimination between .jpgs that contain singing-related spectra and those that do not. Finally, we used a MatLab application (code by L. Cooper, http://www.math.fsu.edu/~bertram/software/birdsong) to convert the file names of remaining false-positive images (i.e., .jpg files that did not contain singing-related spectra) from .jpg back to .wav and to batch-delete those files from the directory of .wav files. Thus, each day of .wav file production by each bird was reduced to a directory that contained only .wav files with song bouts. Although each of these .wav files contained a song bout, some files also included cage noises (pecking on the cage floor, wing flaps, or beak-wiping on the perch) and short and long calls that sometimes occurred in close temporal proximity to the motif.

2.4. Syllable identification

To parse motifs from song bouts into syllable units we used the Feature Batch module in SA+ (version 1.04). Feature Batch generates a spreadsheet representing each syllable and corresponding acoustic characteristics. Although SA+ calculates many acoustic variables, we have determined that a combination of pitch, entropy, pitch goodness, and frequency modulation (FM) are sufficient to effectively individuate zebra finch syllables (for examples, see Fig. 1). Thus, with data from these four features we created 2D scatter plots (each of the four acoustic features vs. syllable duration) to capture syllable structure across multiple bouts.

Fig. 1 illustrates identification of individual syllable clusters from each of the four acoustic feature scatter plots created for a day of preoperative singing. In order to empirically identify individual syllables within 2D scatter plots we use a 'syllable cluster template'.
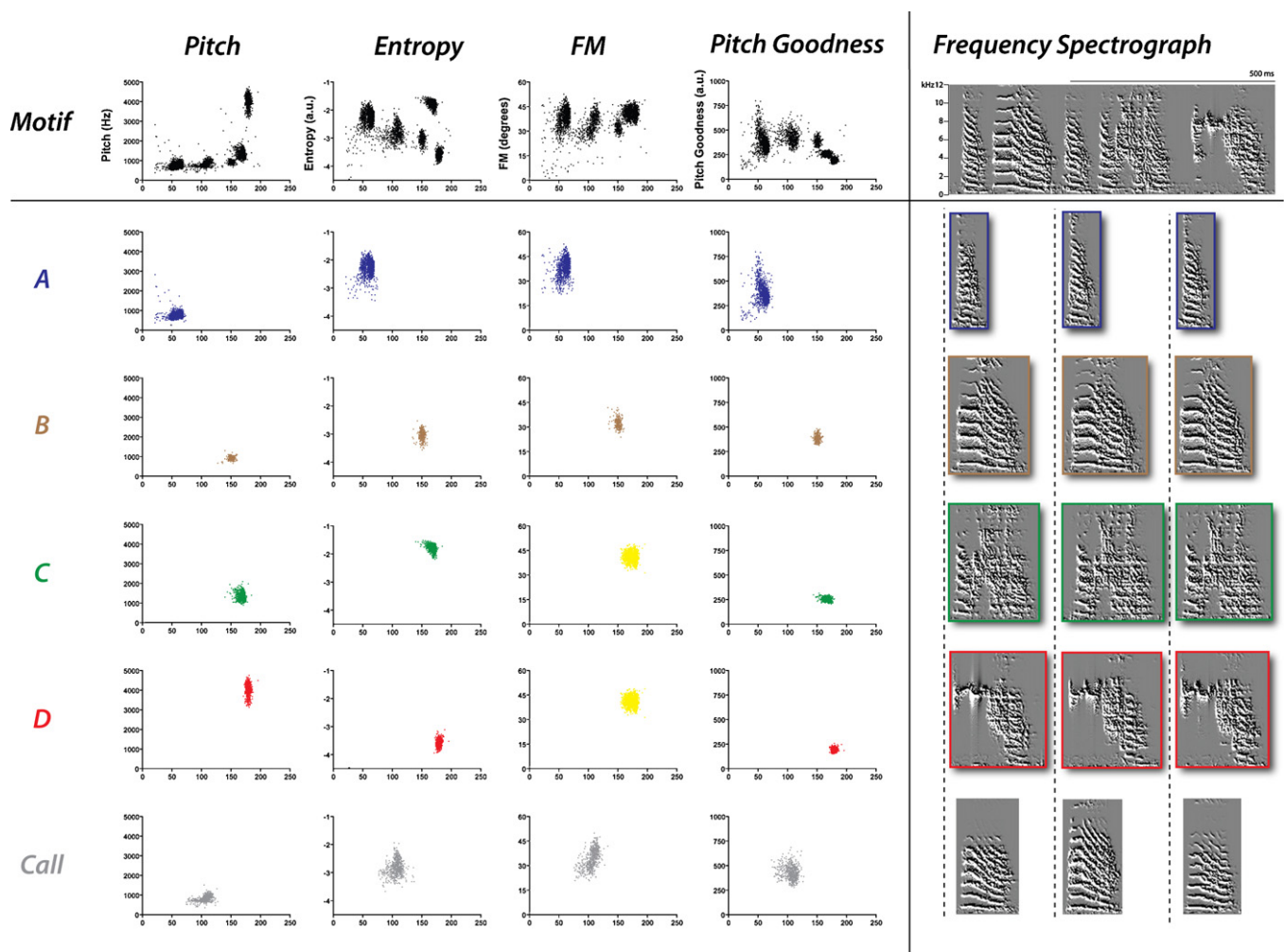
Fig. 1. The columns of the 2D scatter plots represent exemplar data for different acoustic features (pitch, entropy, FM, and pitch goodness) generated from a day of preoperative singing for one bird. The first row of the scatter plots shows the aggregate of all syllables assessed for a day of singing. Each succeeding row shows the acoustic analysis of individual syllables (i.e., syllable A, B, C, D and a call). Although all syllables can be distinguished by at least one feature, syllables that share similar acoustic feature values generate clusters that overlap. For example, syllables C and D produce values that fall within the same cluster when assessed on the basis of FM; yellow clusters indicate undifferentiated syllable clusters. All rows juxtapose exemplar spectrographic representations of measured syllables: one example of the motif and three separate examples of each syllable. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Templates were created by compiling 20 randomly selected examples of each syllable into single .wav files (one .wav file for each syllable). We then applied Feature Batch to each .wav file and generated four scatter plots for each syllable (one for each feature with syllable duration plotted on the x-axis and the acoustic feature plotted on the y-axis); these syllable-specific scatter plots provide a template that defines the acoustic properties of individual syllables. Next, Feature Batch was used to generate four scatter plots from a single day of preoperative song bouts, (see first row of Fig. 1 for exemplar acoustic feature scatter plots). Within these plots, each data point represents an individual syllable and thus discrete clusters of data points signify repeated production of a specific syllable type. We then superimposed the scatter plot for each individual syllable over the scatter plot of all possible syllables which permitted unambiguous partition and identification of individual syllable clusters. At this point it was often necessary to fine-tune Feature Batch threshold settings (amplitude, entropy, minimum syllable duration and minimum gap duration) to achieve an optimal correspondence between syllable clusters generated from the entire day of song bouts and the syllable cluster templates. Final syllable threshold settings were then applied to all days of singing for each bird.

2.5. Two-dimensional statistical characterization of song phonology

As indicated earlier, SA+ Feature Batch calculates the duration, pitch, entropy, pitch goodness, and FM for all selected syllables. Two-dimensional scatter plots can be constructed using the duration and each of the four other features. Scatter plots generated from these acoustic features allow the comparison of large sets of syllables to determine the rate (rapid or gradual) and structure (degradation or cohesion) of changes to the vocal pattern over time. To track changes in the vocal pattern of each bird, feature scatter plots are treated as random samples from a two-dimensional probability distribution and compared to one another; this allows use of the Kullback–Leibler distance, a powerful tool for comparing two probability distributions. As a first step for converting each feature vs. duration scatter plot to a probability density function, the scatter plot is estimated using a nonparametric histogram method. We partition the two-dimensional scatter plot into an M by N array of bins (syllable duration is partitioned into M equally-spaced bins and another feature such as pitch is partitioned into N equally-spaced bins). Thus, there are $M \times N$ two-dimensional bins, as shown in red in Fig. 2A. To have an appropriate and consistent estimation of the
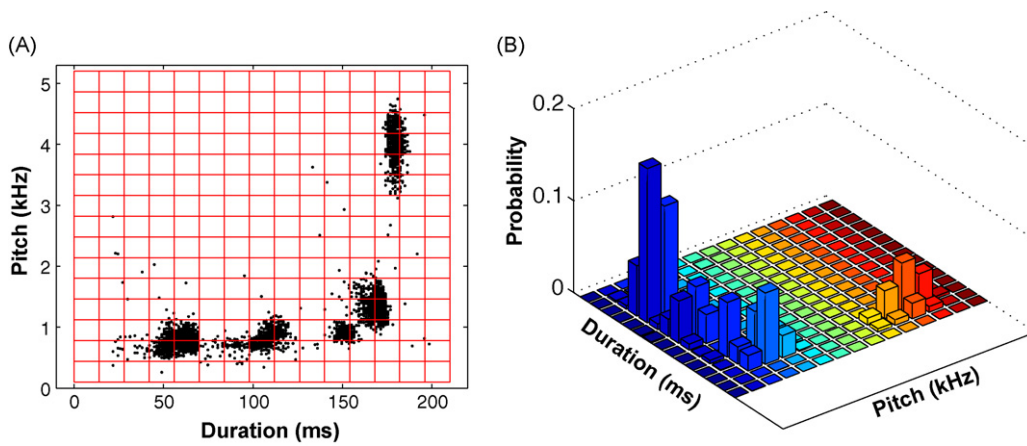
Fig. 2. Example of probability density estimation. (A) Scatter plot of pitch vs. duration for day 1 of singing of a single bird (Bird 611). To use the scatter plot as an estimate of a two-dimensional probability density function, an $M \times N$ grid ($M = N = 15$) is superimposed onto the scatter plot. (B) The estimated probability density function. The height of the rectangular box in each bin denotes its probability. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

distribution, M and N were kept constant (at 15) throughout all days of singing. This number of bins resulted in the best balance between fit to the data and resolution in differentiating syllables. Finally, we counted the number of data points in each bin and divided by the total number of points in the plot. This gave an estimate of the probability density at each location in the plane (Fig. 2B). Note that we use an estimated density function that is based on the raw data. However, a certain degree smoothing may be desirable to better represent the probability distribution, depending on the distribution of the data.

2.6. Scatter plot comparison using the Kullback–Leibler (K–L) distance

The K–L distance analysis allows us to compare the probability density function from two large sets of syllables and quantify the difference. For the behavioral data set included in this example we were interested in observing the rate of recovery or change over time, thus we compared all days of singing to the first baseline day of preoperative singing. We let P_1 represent the two-dimensional scatter plot generated for the first day of singing and P_k the scatter plot for any subsequent day k. Then to compare the difference between P_1 and P_k we compared the estimated probability density functions (denoted Q_1 and Q_k) for the two scatter plots. We then used the K–L distance (Kullback and Leibler, 1951) to compare the density functions. If we let $q_1(m,n)$ and $q_k(m,n)$ denote the estimated probabilities for bin (m,n) for days 1 and k, respectively, then the K–L distance between Q_1 and Q_k is defined as:

$$D_{KL}(Q_1 || Q_k) = \sum_{m=1}^{M} \sum_{n=1}^{N} q_1(m, n) \log_2 \frac{q_1(m, n)}{q_k(m, n)}. \tag{1}$$

Larger values of the K–L distance indicate that two patterns are more dissimilar, and a K–L distance of 0 indicates a perfect match.

2.7. Quantifying the rate of postoperative song recovery (τ)

Given that K–L distance analysis between two probability distributions provides a single number, we plotted the K–L distance value for each pre and postoperative day as a function of time in order to estimate the rate of recovery. We quantified the recovery of the song using an exponential fit to the K–L distance curve after the microlesion. To do this, we first normalized the distance D_{KL} to

the K–L distance at day 4 (the day of the surgery),

$$nD_k = \frac{D_{KL}(Q_1 || Q_k)}{D_{KL}(Q_1 || Q_4)}, \tag{2}$$

where nD_k denotes the normalized K–L distance at the kth session, $k = 4, \ldots, 12$. The normalized K–L distance equals 1 when $k = 4$, then decreases as song recovery occurs. An exponential decay function is then fit to the normalized nD_k curve. That is, we let

$$nD_k = \exp\left(-\frac{k-4}{\tau}\right), \tag{3}$$

where τ denotes the recovery time constant (in units of days). A large τ implies a long-time, or slow, recovery, while a small τ implies a short-time, or fast, recovery.

2.8. Syntax analysis

For each bird, we use the syllable cluster template to identify syllable-types in a scatter plot representing a baseline day of singing (see Fig. 1). Then we trace a polygonal boundary around each cluster to distinguish one syllable from another (Fig. 3). These syllable boundaries are then used to identify syllable types produced on each subsequent day of singing. Any point that falls outside of the syllable boundaries is defined as syllable type 'n', or 'non-motif' (these include occasional calls that occur within a bout). Introductory syllables in a bout are classified as 'intro'.

For the kth day of singing, we count the number of transitions between syllable types i and j, denoted as N_k (i, j). We omit any transition between bouts of singing, counting only syllable transitions within a bout. Spreadsheets generated by SA+ preserve syntatic information for bout and syllable order. Using N_{tot} to denote the total number of transitions made within bouts, the transition probability of syllable syntax (i, j) on day k is estimated as(4)r_k(i, j) = $\frac{N_k(i, j)}{N_{tot}}$ where i and j represent any of the syllable types (as defined on the first day of singing) and $k = 1, 2, \ldots, 12$.

2.9. K–L distance applied to syntax results

As in the phonology analysis, we can calculate the K–L distance between the transition probabilities of two different days to quantify the dissimilarity of song syntax on those days. Since our example involves disruption of the song from and recovery back to the preoperative state, we compared the transition probability distribution of each day to that of the first preoperative day
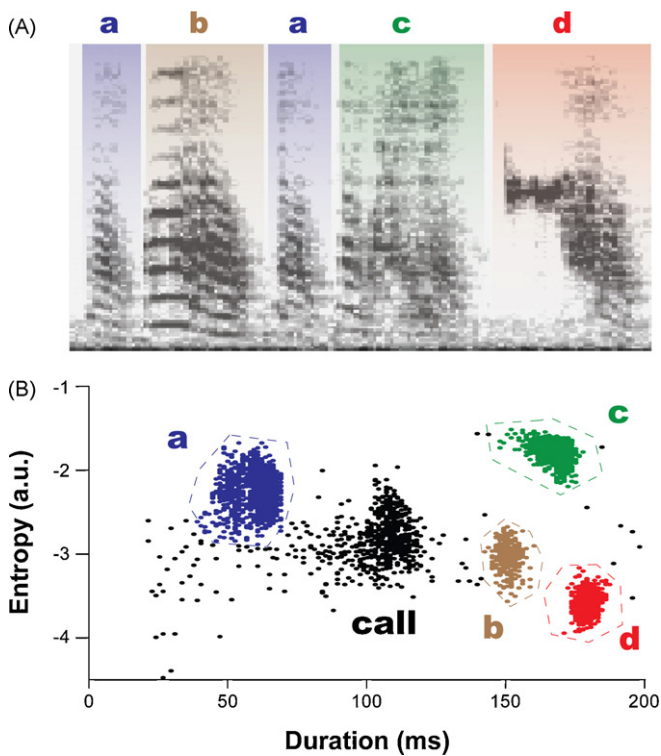
Fig. 3. (A) There are four motif syllables (*a*, *b*, *c*, and *d*) in the song spectrograph in Bird 611. They are denoted by blue, brown, green, and red colors, respectively. (B) Their corresponding syllable clusters in the scatter plot of entropy vs. duration are color-matched to show how syllables were identified for syllable transition probabilities. The syllable types are easily distinguished as individual clusters in the scatter plots. A cluster of black dots outside the syllable boundaries is indicated by "call" and represents an infrequent call produced during singing. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

(day 1). Let R_1, R_2, R_3 denote the estimated transition probability distributions in the three preoperative days, respectively, and R_4, R_5, ..., R_{12} denote the distributions in the nine postoperative days, respectively. For $k = 1, 2, ..., 12$, the K–L distance between R_1 and R_k is:

$$D_{KL}(R_1||R_k) = \sum_{\substack{\text{sequence} \\ (i,j)}} r_1(i, j)\log_2 \frac{r_1(i, j)}{r_k(i, j)} \qquad (5)$$

The summation is over all possible two-syllable transitions (e.g., ab, ac, bc, etc.). Note that because the transition distributions are already between discrete states, there is no need to use a grid to discretize, as was done in the phonology study. The MatLab applications on K–L distance (density estimation, phonology analysis, syntax analysis, and recovery quantification) are available at http://www.math.fsu.edu/~bertram/software/birdsong.

3. Results

To illustrate results obtained by applying our method we use behavioral data from an adult male zebra finch (Bird 611) included in a recent study (Thompson et al., 2007). This bird received bilateral microlesions (see Section 2) targeted to the song region HVC (proper name), a manipulation that produces a transient destabilization of the motif (Thompson and Johnson, 2007; Thompson et al., 2007).

3.1. Scatter plot comparison using the K–L distance

Fig. 4A shows scatter plots for two consecutive preoperative days of singing for Bird 611. The gray dots are from day 1 and the black dots are from day 2. There is a great deal of overlap of the clusters on the two days, indicating that the motifs are similar. The relative similarity in phonology of two preoperative days resulted in a low K–L value of 0.1 bits. In contrast, Fig. 4B shows scatter plots for the preoperative day (gray) and the first day of singing following HVC microlesions (black; resulting in highly destabilized song). The phonological degradation evidenced in day 4 singing bears little resemblance to the clustering pattern of day 1. This is reflected in a very large K–L distance of 4.9 bits. After three additional days (day 7), the match between the pre-lesion and post-lesion scatter plots is much closer, as shown in Fig. 4C with a K–L distance of 1.6 bits. At the final recording session (day 12), the match is almost at the preoperative level with a K–L distance of 0.6 bits (Fig. 4D).

3.2. Quantifying the rate of postoperative song recovery (τ)

Fig. 5A shows an example of vocal change following an HVC microlesion as revealed by K–L distance analysis: days 1–3 are preoperative days of singing, and days 4–12 are days of postoperative singing. For each pair of features (duration vs. pitch, duration vs. entropy, duration vs. pitch goodness, and duration vs. FM), the K–L distance between day 1 and day k is calculated and plotted vs. time, for k = 1, 2, 3, . . ., 12. Preoperative comparisons are highly similar, so the K–L distance between days 1 and days 2 or 3 is near 0. On the first day of singing following surgery (day 4) the K–L distance increases dramatically. On subsequent days of postoperative singing the K–L distance tracks the gradual return in similarity between post HVC microlesion scatter plots and the scatter plot for the first day of preoperative singing.

Fig. 5B shows the recovery quantification of the same data by the commonly used Kolmogorov–Smirnov (K–S) distance (Tchernichovski et al., 2001). Here we calculated the K–S distance between the 1D distribution of each spectral feature (pitch, entropy, pitch goodness, and FM) in day 1 as well as in day k for k = 1, 2, 3, . . ., 12. Consistent with the K–L results, the K–S distances are near 0 at days 2 and 3, increase at day 4, and then gradually decrease during the recovery process. The added constraint of syllable duration in the calculation of K–L distance may account for the clean baseline and clear postoperative trends observed in Fig. 5A.

To assess the rate of recovery for this bird we calculated τ (time constant in Eq. (3)) to measure the recovery rate. For example, we found that τ = 4.0 for the 2D distribution of pitch vs. duration (Fig. 6). Likewise, we found τ = 5.3, 4.5, 6.5, and days for the other three pairs of features (FM vs. duration, entropy vs. duration, and, pitch goodness vs. duration, respectively). These values indicate that the normalized K–L distance fell to 1/e of its original value of 1 in approximately 4–6.5 days, depending on the feature. The same normalization procedure can also be applied to the K–S distance, that is, the recovery time constant τ can be calculated for the K–S values in Fig. 5B. We found that τ = 5.0, 6.8, 5.8, and, 4.5 days for pitch, FM, entropy, and, pitch goodness, respectively. Therefore, K–S distance time constants are comparable to those obtained using K–L distance.

3.3. Syntax analysis

Fig. 7 shows transition probability distributions for Bird 611 over four days. In particular, we refer to the transitions that occur between the syllables in a motif as "motif transitions". The first day represents a preoperative day of singing. There are four different motif syllables in this bird (see Fig. 3). Many of the transitions,
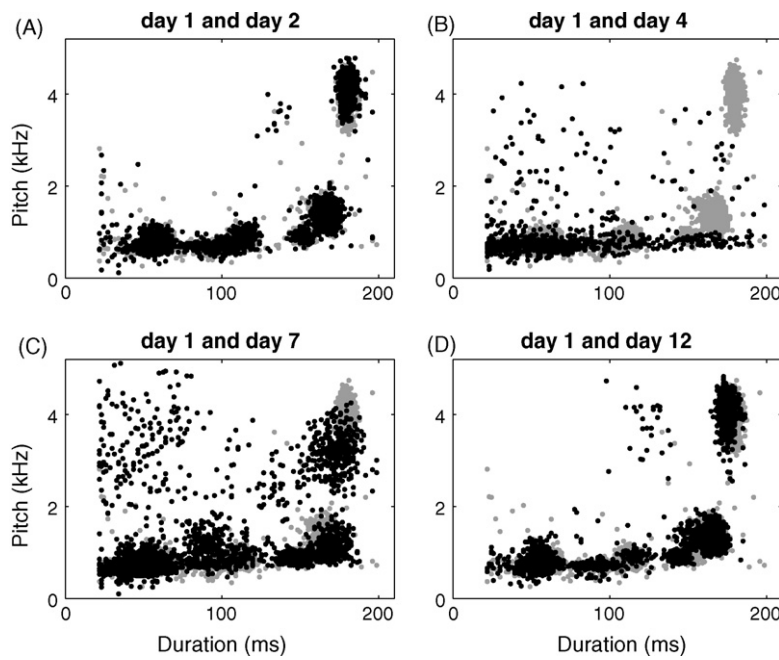
Fig. 4. Examples of K–L distance quantification. (A) Scatter plot of pitch vs. duration for day 1 (Pre1, gray dots) and day 2 (Pre2, black dots) of singing of Bird 611. The K–L distance between the scatter plots is 0.1 bits. (B) Scatter plot of pitch vs. duration for day 1 and day 4 (Post1). The K–L distance is 4.9 bits. (C) Scatter plot of pitch vs. duration for day 1 and day 7 (Post4). The K–L distance is 1.6 bits. (D) Scatter plot of pitch vs. duration for day 1 and day 12 (Post9). The K–L distance is 0.6 bits.

such as an a to d transition, occur only rarely or not at all, so in the figure we grouped all such transitions (with probability less than 0.05 in each of the recording days) into one category called 'others'. Preoperative day 1 shows a high probability for motif syllable transitions and a low probability for non-motif syllable transitions. In contrast, the first day of singing following HVC microlesions (day 4) shows the exact reverse, with a low probability for motif transitions and higher probability for transitions associated with non-motif syllables or syllables that fall outside the boundaries of the syllable clusters. A later postoperative day of singing (day 6) shows some recovery of motif syllable transitions, yet there remains

an increased probability for transitions associated with non-motif syllables. At day 8, one week postoperative, the overall pattern of syllable transition probabilities appears similar to preoperative syntax structure: high probability for motif syllable transitions and low probability for transitions associated with other syllables and non-motif syllables.

3.4. K–L distance applied to syntax results

Eq. (5) was used to compute the K–L distance between the transition probability distribution of day 1 with those of subsequent days. Fig. 8 shows that on days 2 and 3 (additional preoperative days) the song syntax is very similar to that of day 1, so the K–L distance is near 0. On day 4 (first day of singing following the microlesion) the distance increases dramatically to near 8 bits, but then decreases as the song syntax recovers. Thus, the return of the song syntax that is apparent in Fig. 7 is clearly apparent in the time course of the K–L distance in Fig. 8. We also assessed the rate of syntax recovery using τ and found that $\tau = 1.8$ days.
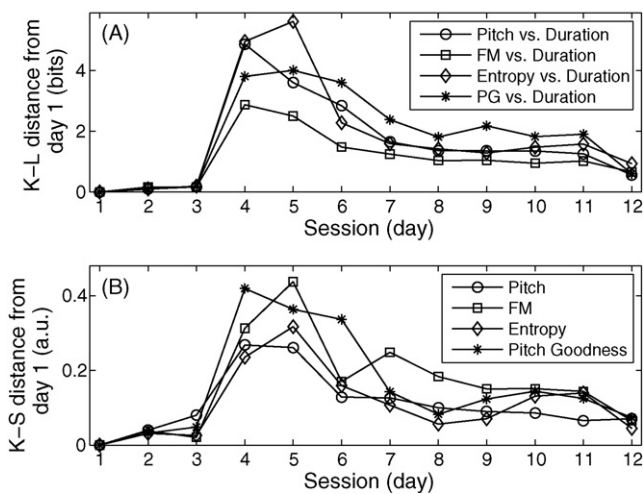

Fig. 5. Comparison of K–L distance and K–S distance values for phonology. (A) K–L distances of each spectral feature vs. duration from day 1 for Bird 611. These values describe a marked increase in phonological dissimilarity from day 1 (at day 4) that gradually declines during the recovery process (day 4 to day 12). (B) K–S distances of each spectral feature from day 1 for Bird 611. These values also describe an increase in dissimilarity at day 4 that gradually declines during the recovery process.
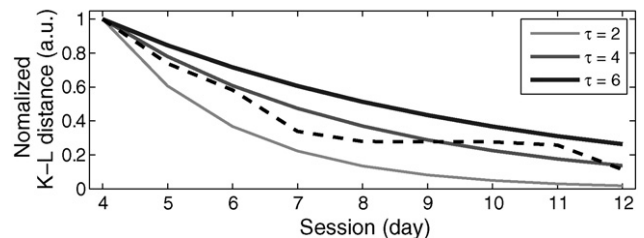

Fig. 6. Song recovery following a perturbing event such as HVC microlesion can be characterized by a recovery time constant τ. The dashed line shows the K–L distance values in the recovery process (day 4 to day 12) of Bird 611, and the solid line in the center denotes their optimal exponential fit with $\tau = 4$. The other two solid lines denote exponential decay with other values of τ for comparison. The upper one indicates a slower recovery ($\tau = 6$), and the lower one indicates a faster recovery ($\tau = 2$).
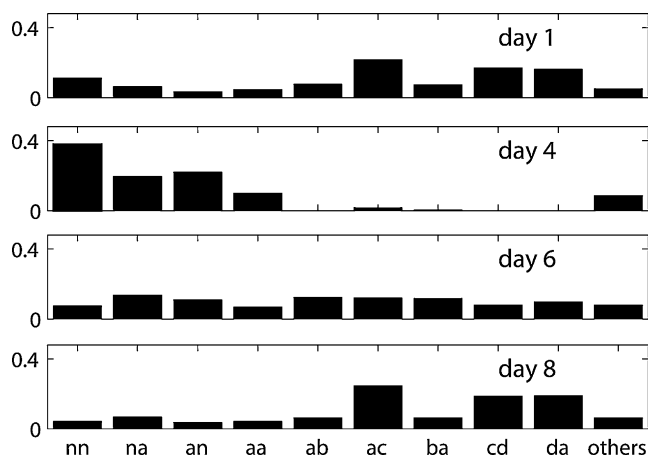
Fig. 7. Syllable transition probability for four days (1, 4, 6, and 8) of Bird 611. The motif transitions are *aa*, *ab*, *ac*, *ba*, *cd*, *da*, which nearly disappeared right after the HVC surgery (day 4), then gradually recovered in the following days. The sums of all motif transitions on each day are 0.74, 0.12, 0.24, and 0.60, respectively.

To determine how production of non-motif syllable transitions affected the rate of recovery, we also characterized syntax recovery excluding non-motif syllable transitions (i.e., 'n'; non-motif). That is, we computed the K–L distance using only the transitions of syllables that fell within the boundaries traced around each motif syllable cluster. Excluding non-motif syllable transitions resulted in a τ of 1.5 days (Fig. 8), which indicates a slightly faster recovery than when non-motif syllable transitions were included (where $\tau = 1.8$ days). This analysis shows that recovery of motif syllable syntax occurred slightly before the production of non-motif syllable transitions had returned to baseline levels.

4. Discussion

We have presented a method for analyzing birdsong phonology and syllable syntax measured with the Feature Batch module in Sound Analysis Pro (SA+; Tchernichovski et al., 2000). This method compares motifs at the syllable-level using a standard measurement from information theory, the Kullback–Leibler distance (Kullback and Leibler, 1951). We have shown that the new method successfully describes differences in multi-dimensional syllable features using dissimilarity values (the K–L distance in units of bits) for motif phonology and syntax, respectively.

The Similarity Batch function in SA+ is a commonly used method in birdsong analysis to measure song similarity. For example, the method has been used to assess the vocal imitation of pupils from tutors (Tchernichovski et al., 2000) or recovery of song following brain injury (Coleman and Vu, 2005; Thompson and Johnson,
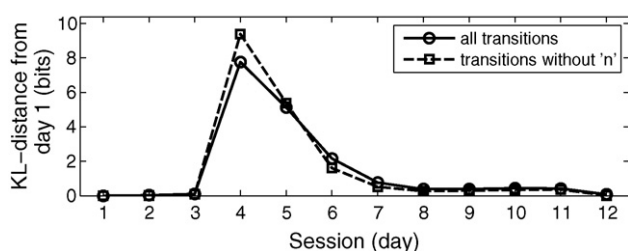

Fig. 8. K–L distance values for syllable syntax. The solid line with circles shows the distances using all transitions. The dashed line with squares shows the distances using transitions that do not include non-motif syllable transitions 'n'. These values describe a marked increase in syllable syntax dissimilarity from day 1 (at day 4) that declines during the recovery process (day 4 to day 12).

2007). The Similarity Batch function is typically used to search for similarity between a single 'target' motif and a .wav file comprised of multiple motifs or a set of uncategorized song units (e.g. destabilized singing following HVC microlesions). This motif-based comparison does not determine syllable-level contributions to similarity. In contrast, the K–L distance is based on the distributions of acoustic features of individual syllables or of their temporal sequence, and thus, dissimilarity between two probability distributions can be traced back to individual syllable contributions. Moreover, by examining the scatter plots used to generate K–L distance values one can determine whether dissimilarity is due to increased variability in the phonology of motif syllables (change in size or shape of syllable clusters), the production of non-motif syllables (syllables that fall outside of motif polygonal clusters), or some combination of the two.

An alternative statistical method to the K–L distance is the Kolmogorov–Smirnov (K–S) test. This is a classical statistical hypothesis test for comparing distributions of two samples. The associated K–S distance, the largest distance between two cumulative distribution functions, is a key component of acoustic analyses performed by Sound Analysis Pro (Tchernichovski et al., 2001). Indeed, we have shown that K–S distance and K–L distance produce similar results for the recovery of each acoustic feature. However, in practice K–S distance, which measures dissimilarity of cumulative distribution functions, is generally limited to one-dimensional quantitative variables. This limitation tends to preclude differentiation between all syllables within the data set for any given acoustic feature. For example, in Fig. 1 most syllables overlap on the *y*-axis (spectral feature dimension); however the addition of duration (*x*-axis) to pitch, entropy, and pitch goodness allows clear separation and differentiation between all syllables, which is necessary for our method of syntax analysis.

In addition, unlike K–L distance, K–S distance cannot be applied to categorical variables. For example, in this study we used K–L distance to measure phonology (two-dimensional distributions of quantitative spectral features) as well as the recovery of syllable syntax, which identifies each cluster with a syllable label and is thus a categorical variable. When τ was calculated for the K–L distance values for syntax and phonology we found that the two had different rates of recovery (i.e., ~5 days for phonology when all acoustic features are averaged and ~1.5 days for syntax). Therefore, K–L distance provides a unified measurement strategy for both phonology and syntax.

4.1. Limitations and considerations

The K–L distance measurement is based on estimated statistical distributions, and a large number of syllables are therefore needed in the density estimation. The method would not be appropriate for situations where only a few samples of singing are available for comparison. For example, the data presented above were generated from the first 300 song bouts that Bird 611 produced each day, resulting in the analysis of ~3500 syllables per day. However, we have found that the method remains surprisingly robust when substantially smaller samples are used – a reanalysis of Bird 611 using 350 syllables per day (a sub-sample of one-tenth the size of the original sample of syllables) still provided a reasonable estimation of the recovery of phonology ($\tau = 4.7$ days vs. $\tau = 4.0$ days from the original sample).

Investigators using our method must also decide on the number of bins aligned against each axis of the 2D scatter plots, which effectively determines the precision of the analysis. We have used a constant, 15, for all the comparisons. However, we find that other constants (varying from 10 to 20) produce similar results. Investigators should choose a reasonable number based on the range

of the data, pattern of the distribution, and the number of sample points. In order to employ the syntax analysis, the investigator must determine the number of syllable clusters unique to an individual bird. We present an efficient approach to classify syllable clusters using well-separated polygons, in which repeated instances of the same syllable are easily identified (Fig. 1). Other methods are available to identify syllable clusters, such as the Clustering Module in SA+ (Tchernichovski et al., 2000), and the KlustaKwik algorithm (Crandall et al., 2007) and the K–L distance on the syntax can be applied to any of these methods.

Acknowledgements

References

Ashmore RC, Bourjaily M, Schmidt MF. Hemispheric coordination is necessary for song production in adult birds: implications for a dual role for forebrain nuclei in vocal motor control. J Neurophysiol 2008;99:373–85.

Cardin JA, Raksin JN, Schmidt MF. Sensorimotor nucleus NIf is necessary for auditory processing but not vocal motor output in the avian song system. J Neurophysiol 2005;93:2157–66.

Coleman MJ, Vu ET. Recovery of impaired songs following unilateral but not bilateral lesions of nucleus uvaeformis of adult zebra finches. J Neurobiol 2005;63: 70–89.

Crandall SR, Aoki N, Nick TA. Developmental modulation of the temporal relationship between brain and behavior. J Neurophysiol 2007;97:806–16.

Deregnaucourt S, Mitra PP, Feher O, Pytte C, Tchernichovski O. How sleep affects the developmental learning of bird song. Nature 2005;433:710–6.

Funabiki Y, Konishi M. Long memory in song learning by zebra finches. J Neurosci 2003;23:6928–35.

Haesler S, Rochefort C, Georgi B, Licznerski P, Osten P, Scharff C. Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. PLoS Biol 2007;5:e321.

Hara E, Kubikova L, Hessler NA, Jarvis ED. Role of the midbrain dopaminergic system in modulation of vocal brain activation by social context. Eur J Neurosci 2007;25:3406–16.

Heinrich JE, Nordeen KW, Nordeen EJ. Dissociation between extension of the sensitive period for avian vocal learning and dendritic spine loss in the song nucleus lMAN. Neurobiol Learn Mem 2005;83:143–50.

Hough 2nd GE, Volman SF. Short-term and long-term effects of vocal distortion on song maintenance in zebra finches. J Neurosci 2002;22:1177–86.

Kittelberger JM, Mooney R. Acute injections of brain-derived neurotrophic factor in a vocal premotor nucleus reversibly disrupt adult birdsong stability and trigger syllable deletion. J Neurobiol 2005;62:406–24.

Kullback S, Leibler RA. On information and sufficiency. Ann Math Stat 1951;22:79–86.

Liu WC, Nottebohm F. Variable rate of singing and variable song duration are associated with high immediate early gene expression in two anterior forebrain song nuclei. Proc Natl Acad Sci U S A 2005;102:10724–9.

Liu WC, Nottebohm F. A learning program that ensures prompt and versatile vocal imitation. Proc Natl Acad Sci U S A 2007;104:20398–403.

Liu WC, Gardner TJ, Nottebohm F. Juvenile zebra finches can use multiple strategies to learn the same song. Proc Natl Acad Sci U S A 2004;101:18177–82.

Olveczky BP, Andalman AS, Fee MS. Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. PLoS Biol 2005;3:e153.

Phan ML, Pytte CL, Vicario DS. Early auditory experience generates long-lasting memories that may subserve vocal learning in songbirds. Proc Natl Acad Sci U S A 2006;103:1088–93.

Pytte CL, Gerson M, Miller J, Kirn JR. Increasing stereotypy in adult zebra finch song correlates with a declining rate of adult neurogenesis. Dev Neurobiol 2007;67:1699–720.

Roy A, Mooney R. Auditory plasticity in a basal ganglia-forebrain pathway during decrystallization of adult birdsong. J Neurosci 2007;27:6374–87.

Sakata JT, Brainard MS. Real-time contributions of auditory feedback to avian vocal motor control. J Neurosci 2006;26:9619–28.

Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, Mitra PP. A procedure for an automated measurement of song similarity. Anim Behav 2000;59:1167–76.

Tchernichovski O, Mitra PP, Lints T, Nottebohm F. Dynamics of the vocal imitation process: how a zebra finch learns its song. Science 2001;291:2564–9.

Teramitsu I, White SA. FoxP2 regulation during undirected singing in adult songbirds. J Neurosci 2006;26:7390–4.

Thompson JA, Johnson F. HVC microlesions do not destabilize the vocal patterns of adult male zebra finches with prior ablation of LMAN. Dev Neurobiol 2007;67:205–18.

Thompson JA, Wu W, Bertram R, Johnson F. Auditory-dependent vocal recovery in adult male zebra finches is facilitated by lesion of a forebrain pathway that includes the basal ganglia. J Neurosci 2007;27:12308–20.

Wilbrecht L, Crionas A, Nottebohm F. Experience affects recruitment of new neurons but not adult neuron number. J Neurosci 2002a;22:825–31.

Wilbrecht L, Petersen T, Nottebohm F. Bilateral LMAN lesions cancel differences in HVC neuronal recruitment induced by unilateral syringeal denervation. Lateral magnocellular nucleus of the anterior neostriatum. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 2002b;188:909–15.

Wilbrecht L, Williams H, Gangadhar N, Nottebohm F. High levels of new neuron addition persist when the sensitive period for song learning is experimentally prolonged. J Neurosci 2006;26:9135–41.

Woolley SM, Rubel EW. Vocal memory and learning in adult Bengalese Finches with regenerated hair cells. J Neurosci 2002;22:7774–87.

Zann R, Cash E. Developmental stress impairs song complexity but not learning accuracy in non-domesticated zebra finches (*Taeniopygia guttata*). Behav Ecol Sociobiol 2008;62:391–400.

Zevin JD, Seidenberg MS, Bottjer SW. Limits on reacquisition of song in adult zebra finches exposed to white noise. J Neurosci 2004;24:5849–62.