

Basic Neuroscience

A computational tool for automated large-scale analysis and measurement of bird-song syntax

Arij Daou^a, Frank Johnson^{b,c}, Wei Wu^{c,d}, Richard Bertram^{a,c,*}^a Department of Mathematics, Florida State University, United States^b Department of Psychology, Florida State University, United States^c Program in Neuroscience, Florida State University, United States^d Department of Statistics, Florida State University, United States

HIGHLIGHTS

- ▶ We describe new computer software for the automated analysis of the syntactic structure of birdsong. This replaces the time-consuming manual analysis that is commonly done.
- ▶ Application includes analysis of bird vocalizations as the bird matures and the song progresses to the adult song.
- ▶ The software can also be applied to the analysis of adult birdsong and the effects of neural perturbations or external factors.
- ▶ Several analysis tools are employed, including linearity, consistency, and stereotypy scores.

ARTICLE INFO

Article history:

Received 1 February 2012

Received in revised form 23 July 2012

Accepted 23 July 2012

Keywords:

Birdsong
 Sequential behavior
 Behavioral analysis
 Computer software

ABSTRACT

We present computer software for automated, high throughput, quantitative syllable-level analysis of bird song syntax. The primary advantage of our tool is the ease and effectiveness it provides in quantifying syllable sequence and performing a comparison of syllable sequence from one day of singing with one or more other days of singing. The software utilizes the output of the Feature Batch module in Sound Analysis Pro (Tchernichovski et al., 2000) that can be used to measure the temporal and spectral features of each syllable produced during a day of singing. We use these measurements to identify individual syllables based on their temporal and spectral properties and then identify transition probabilities among syllables to determine changes in syntax. This quantifies the ordering of syllables in songs and the frequency with which subsequences appear. Moreover, the software computes the linearity, consistency, and stereotypy scores for every bout presented as well as descriptive statistics for each of these measures for each day of singing. We also report statistical measures that the software utilizes (the Kullback–Leibler distance and the sequence entropy) to quantify the degree of dissimilarity between sequences of syllable transitions. Our tool is useful for comparing the syntactic structure of songs produced by a bird prior to and after a manipulation such as ablation of part of the vocal motor pathway or infusion of pharmacological agents, or for assessing the degree of individual variation in syntactic structure across populations of birds.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Songbirds such as zebra finches and Bengalese finches are often studied due to their patterned song. One important feature of a song is the sequence of syllables that comprise it. Here we describe a computer software tool, *SongSeq*, for the analysis of the transitions made in one or more bouts of singing by a bird, using the zebra finch as a model system.

The input to *SongSeq* is the output from the Sound Analysis Pro (SA+) software package. SA+ is a frequently used software package for converting continuous sound into discrete syllables and extracting features such as duration, pitch, pitch goodness and entropy from the syllables (Tchernichovski et al., 2000).

SongSeq can be used to monitor changes of sound features across a large number of songs, analyze transition probabilities among syllables, quantify the consistency of syllable ordering (linearity, consistency and stereotypy scores), and quantify the degree of similarity in syllable phonology over different days of singing. In order to generate syllable-level comparisons, we developed a graphical user interface (GUI) to manually identify syllable clusters (i.e., repeated instances of the same syllable). Other methods to identify

* Corresponding author at. Department of Mathematics, Florida State University, Tallahassee, FL 32306-4510, United States. Tel.: +1 850 644 7195.

E-mail address: bertram@math.fsu.edu (R. Bertram).

syllable clusters are available – for example, the Clustering Module in SA+ which identifies individual syllables based on Euclidean distance and high dimension filtering in acoustic feature spaces. The method presented here can be applied regardless of the syllable cluster strategy employed.

SongSeq is an easy and effective tool for quantifying the development of syllable sequences and performing a multi-dimensional comparison of the syllable acoustic features from one day of singing with songs from other days. This would be a useful tool for comparing the songs produced by a bird at different stages of development or prior to and after a manipulation such as partial brain ablation to portion(s) of the neural song system or infusion of pharmacological agents.

2. Methods

The software described here is available for free download from <http://www.math.fsu.edu/~bertram/software/songbird>. There is also an online tutorial and a user manual at this site.

Recordings of bird song are read into the SA+ software package as .wav files. Syllable units are generated by parsing every motif from song bouts using syllable segmentation tools in the Feature Batch module in SA+. Feature Batch generates a spreadsheet for all syllables and their acoustic characteristics (for further details, see Wu et al., 2008). These spreadsheets are the input to *SongSeq*.

The entire procedure that *SongSeq* goes through is discussed below in detail and can be summarized by the following brief overview: (1) choose a template SAP spreadsheet, (2) choose two acoustic features to use for identifying syllable clusters, (3) name the syllables and assign them different colors, (4) use colored boxes to paint the different syllable clusters, (5) if syllables do not cluster unambiguously, repeat steps 2 and 4 on two different acoustic features, (6) choose the test song files, and (7) compare features of test songs to those from the template.

2.1. Identification of syllables using a template

To compare syllable features and sequences between one singing session and another, one must first designate one spreadsheet as the template (a typical SA+ spreadsheet containing the acoustic variables, which could be the data points for a day of pre-operative singing, for example). Two acoustic features are then selected by the user from the template using *SongSeq*'s graphical user interface (Fig. 1). These features are extracted from each syllable in the template sonogram (Fig. 2A) and displayed as a 2D scatter plot (Fig. 2B). Within the scatter plot, each data point represents an instance of a syllable and discrete clusters of data points signify repeated production of a specific syllable type. This captures the syllable structure across multiple bouts, and defines the acoustic properties of individual syllables. The choice of acoustic features used to form the scatter plot could be important, since some feature pairs may be more effective than others in discriminating syllables. One strategy is to just try different combinations; *SongSeq* contains a module that enables the user to browse the different features and choose the best two features for syllable discrimination. An initial eye-identification of the spectrograms is helpful where the number of different syllables can be identified along with other acoustic features of the various syllables (like syllable duration, mean FM, amplitude and entropy). The number of clusters on the scatter plot should equal the number of different syllables on the spectrogram. We have found that syllable duration is typically a good feature to use (Wu et al., 2008). If two features are not enough to unambiguously discriminate features, *SongSeq* allows one to use a second pair of features for further discrimination.

The next step in syllable identification is using the graphical tools to associate syllable clusters with syllable labels (A, B, C, etc.) and colors. This is done using mouse clicks over the area in the 2D scatter plot that represents the syllable. Each mouse click paints onto the scatter plot a colored box covering the instances of a syllable (e.g., syllable A). This is used to define the boundary of

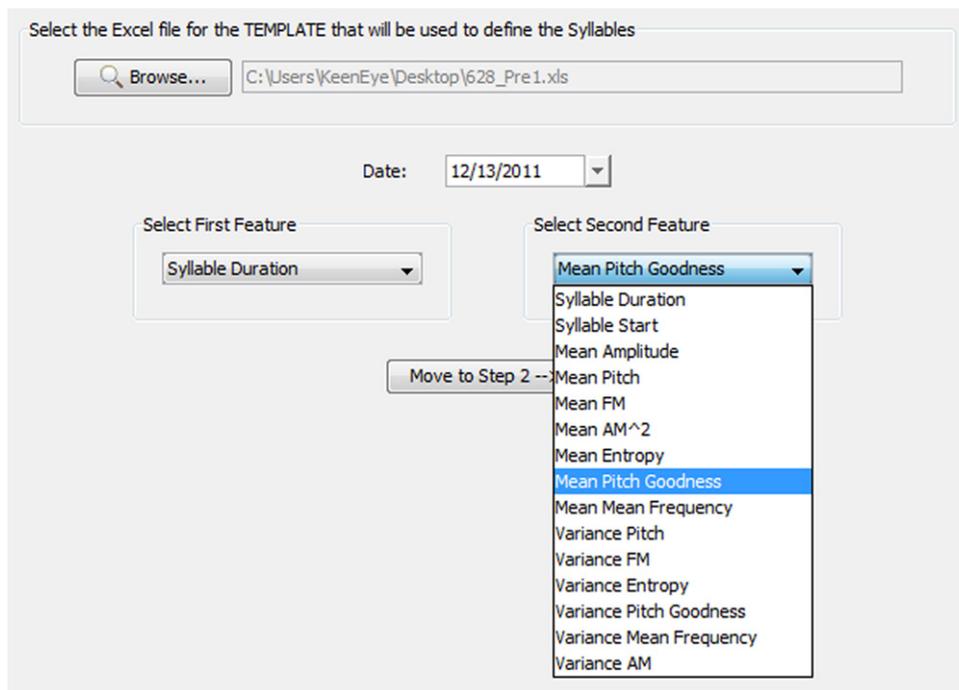


Fig. 1. Designation of the template spreadsheet and selection of two acoustic features. The user browses and selects a template file (SA+ spreadsheet). The template's file storage path is displayed in the textbox after selection. Two acoustic features are then selected by the user. The acoustic features drop down menu lists all the features that SA+ generates.

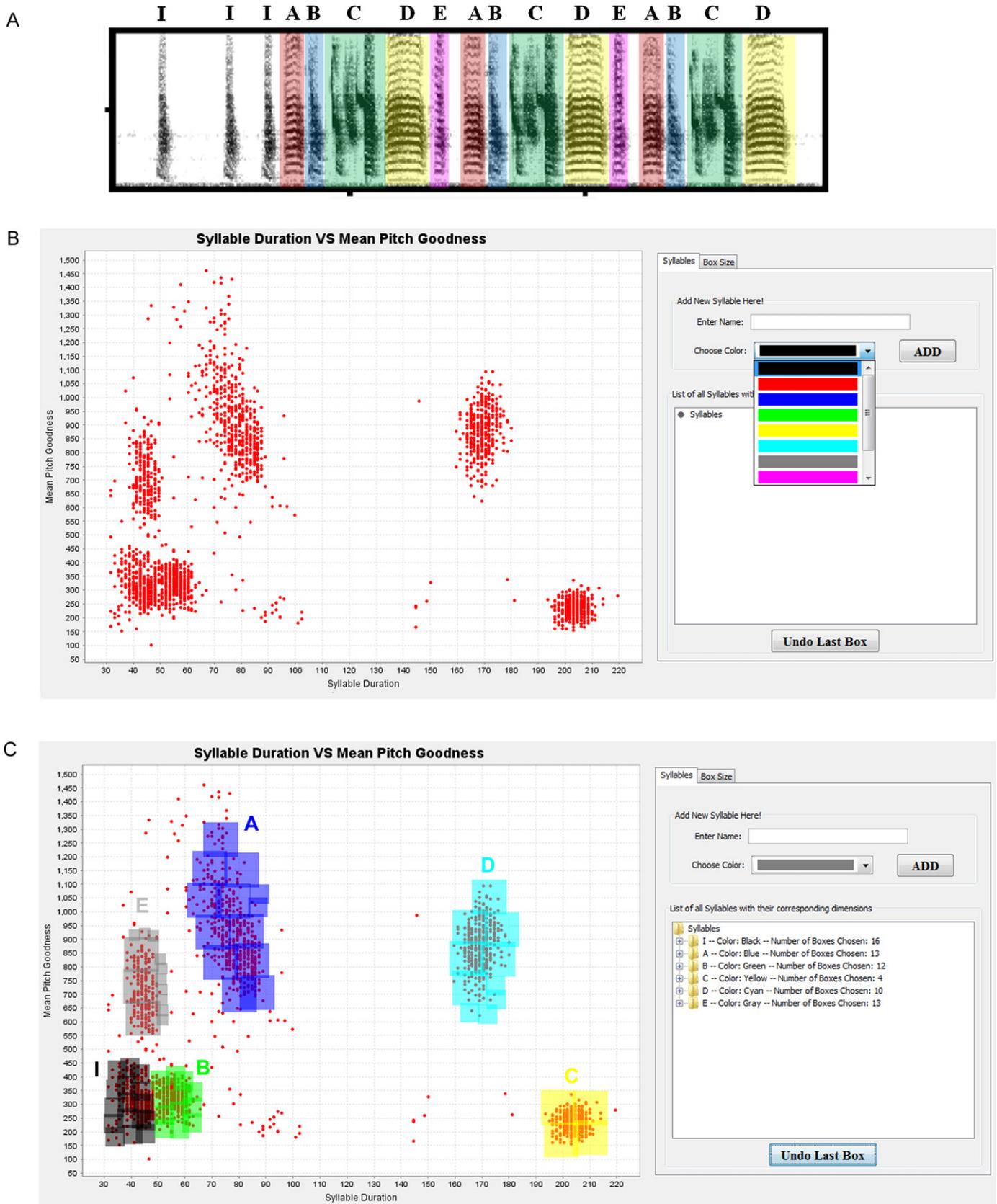


Fig. 2. Syllable identification. (A) Sonogram of a typical song of a bird that received bilateral HVC microlesions displaying three introductory notes (labeled as I) and five syllables (labeled as A, B, C, D, and E). (B) A screen shot of *SongSeq*'s frame where syllable identification is processed. *Left panel*: a 2D scatter plot of the two selected features from a preoperative day of singing (syllable duration versus mean pitch goodness). *Right panel*: GUI that associates the syllables clusters on the 2D scatter plot with syllable labels and colors. (C) Using mouse clicks, the instances of each syllable in (B) are painted onto the scatter plot. The region covered by boxes of a single color defines a syllable. Here we have five syllables along with a cluster of introductory notes that is painted in black as I. On the right hand side of the frame, a tree shows the number of boxes used for each syllable. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

that syllable in subsequent scatter plots (e.g., for different days of singing). The region covered by boxes of a single color defines a syllable (e.g., Fig. 2C shows five syllables labeled as A, B, C, D and E along with a cluster of introductory notes labeled as I). The colored boxes can be resized by the user (e.g., syllables A and E in Fig. 2C are painted with boxes of different sizes). The boxes are translucent so that one can see the syllable instances (dots) under the clusters. This procedure of syllable identification is done for every syllable, and thus for every cluster.

It may happen that some syllables are easily identified with one pair of features, while others are better identified with a second pair of features. With *SongSeq*, one can identify the first set of syllables using one feature pair, then move to a second feature pair to improve the identification of the remaining syllables. In this process, and after all the syllables are identified with the first 2 acoustic features (Fig. 2C), two new acoustic features are chosen along with a subset of the previously added syllables. Only this chosen subset of syllables (which can include all syllables) can be painted again in the new 2D feature space. Fig. 3 shows the selection of syllable duration versus mean entropy acoustic features. It also shows the selection of B and I since their clusters in Fig. 2C are not well isolated, and as we will see shortly syllable duration versus mean entropy pulls B and I apart. Syllables A, C, D and E are not chosen here because they are discriminated nicely with the first two acoustic features (Fig. 2C).

Once the second pair of features is chosen, a new 2D scatter plot appears. The data points are now color coded according to the first step of identification (Fig. 4A). For instance, any data points that were within the blue boxes defining the borders of syllable A in the scatter plot of Fig. 2C, will be painted in the same color (blue) in the scatter plot of Fig. 4A. In Fig. 4A we see that the data points for B (green) and I (black) are pulled apart in the new 2D feature space, and it is clear that some data points previously misidentified as B actually cluster better with I (green points in the bottom right of the black cluster). Syllables A (blue), C (yellow) and D (cyan) remain well isolated, but syllable E (gray) has a very similar mean entropy as the introductory notes and therefore their corresponding clusters overlap.

Next, the user paints in the same way as before onto the new scatter plot, but now only for the syllables (B and I) that were

specified for this second round of discrimination (Fig. 4B). For these specified syllables, a data point is considered as syllable X if and only if either of the following two conditions hold: (1) it belongs to one of the boxes defining syllable X in the first 2D scatter plot AND it belongs to one of the boxes defining syllable X in the second 2D scatter plot, (2) it belongs to one of the boxes defining syllable X in the second 2D scatter plot BUT is not a data point of a non-specified syllable (that is, a data point that is colored with blue (A), yellow (C), cyan (D) or gray (E)). Although the clusters for E and I overlap in the second scatter plot, the points originally labeled as E retain that identification since E was not selected for repainting.

The next step is choosing the target files. Target files are typical spreadsheets generated by SA+. Information from these target files (which could be day(s) of post-operative singing, as in Fig. 5) is extracted and compared with the template spreadsheet to obtain sequencing details. Finally, typical transitions are selected by the user. A typical transition between two syllables is defined as one that is frequently encountered. As we will see, the typical transitions are used to calculate the consistency score that reflects the frequency with which a main or typical sequence appears.

2.2. Sequencing the data of the target files

After *SongSeq* receives all required input, it probes the target files in the order they were uploaded. The operations described next are done automatically by *SongSeq* for each target data file. First, rows are associated with syllables based on the values of the user-specified features and are copied into a new spreadsheet, one for each syllable. This procedure is done for every defined syllable. Data points that do not fall into any of the named syllables (e.g., the points not lying in any of the colored regions in Figs. 2C and 4A) are inserted into a spreadsheet called “NMS” (Non-Motif Syllable). Thus, each target file is parsed into a set of spreadsheets: one spreadsheet for every syllable name along with a spreadsheet for “NMS”. Next, these syllable spreadsheets are merged into a final “sequenced” spreadsheet that is created to contain all the notes in the order they were sung. This allows for the analysis of syllable transition probabilities. The different epochs of singing are identified in this spreadsheet by the .wav identifier.

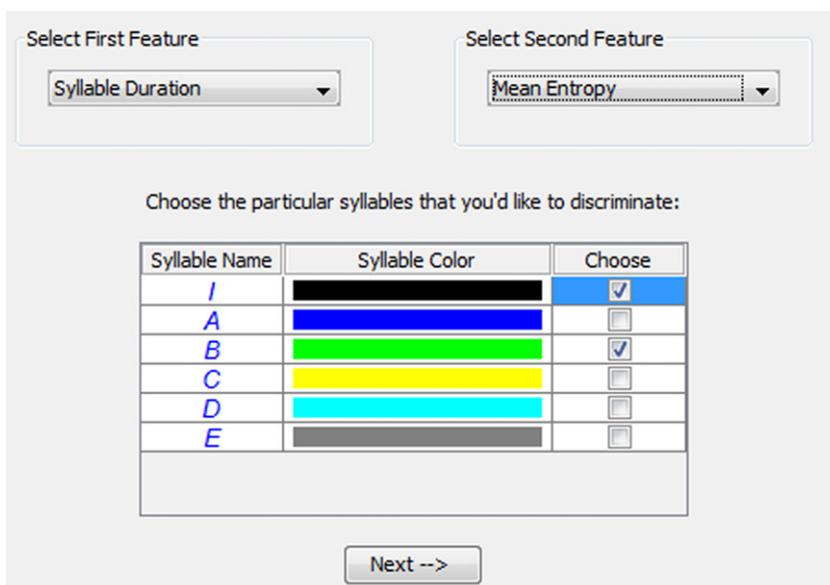


Fig. 3. A screenshot of *SongSeq*'s frame showing the selection of the second pair of acoustic features, and the selection of B and I to better discriminate them on the scatter plot of the new pair of features.

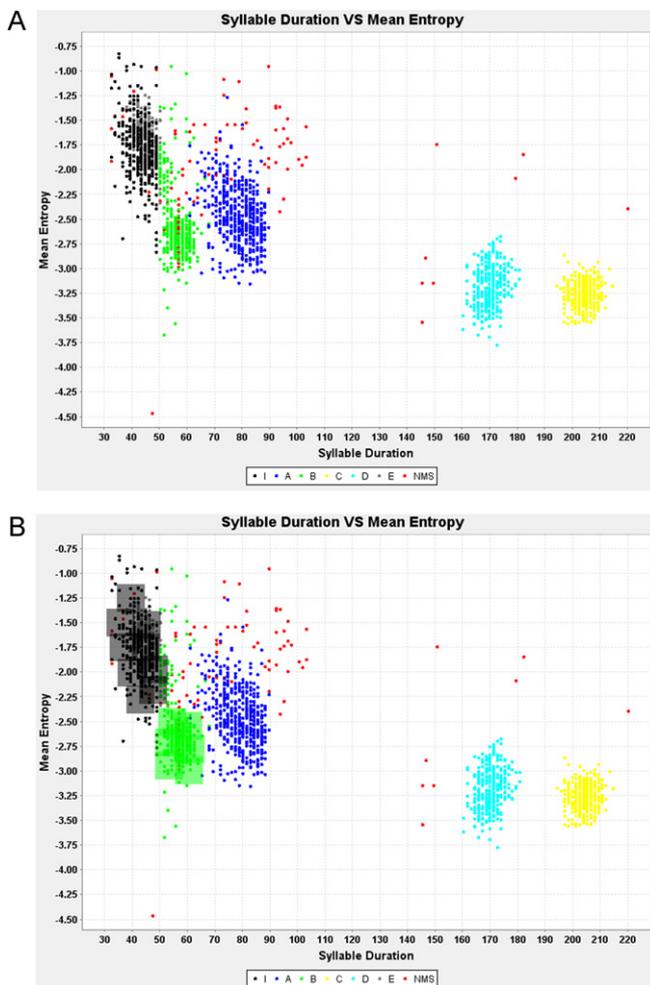


Fig. 4. Second step of syllable identification. (A) Color-coded data points on the scatter plot of the second pair of acoustic features chosen in Fig. 3, where colors are coded based on the painting done in the first step of identification. It is clear that some data points for B (green) and I (black) are pulled apart in the new 2D feature space, and some data points previously misidentified as B actually cluster better with I (green points in the bottom right of the black I cluster). (B) Syllables B and I are painted in the new feature scatter plot. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

2.3. Syntax identification

Syllable transitions and scores are generated next. The sequenced spreadsheet is scanned row by row. For every block of consecutive rows that have the same .wav identifier, syllable transitions are determined. If the syllable name in row i of a song bout is "A" and the syllable name in row $i + 1$ of the same song bout is "B", then the syllable transition is "A–B". This is done for each pair of consecutive rows within the block, producing a list of syllable transitions for every song bout. The syllable transitions are then merged and the number of occurrences of every transition is calculated. A transition probability is calculated by dividing the number of occurrences of the syllable transition by the total number of syllable transitions. The transition probabilities along with the syllable transition name are then written into a new spreadsheet.

The song stereotypy is quantified using two measures that address related but different aspects of sequence stereotypy: *sequence linearity* addresses the way syllables are ordered in a song, and *sequence consistency* addresses how often a particular path is actually followed (Scharff and Nottebohm, 1991). *SongSeq* calculates the linearity score of every song bout (.wav file) by dividing

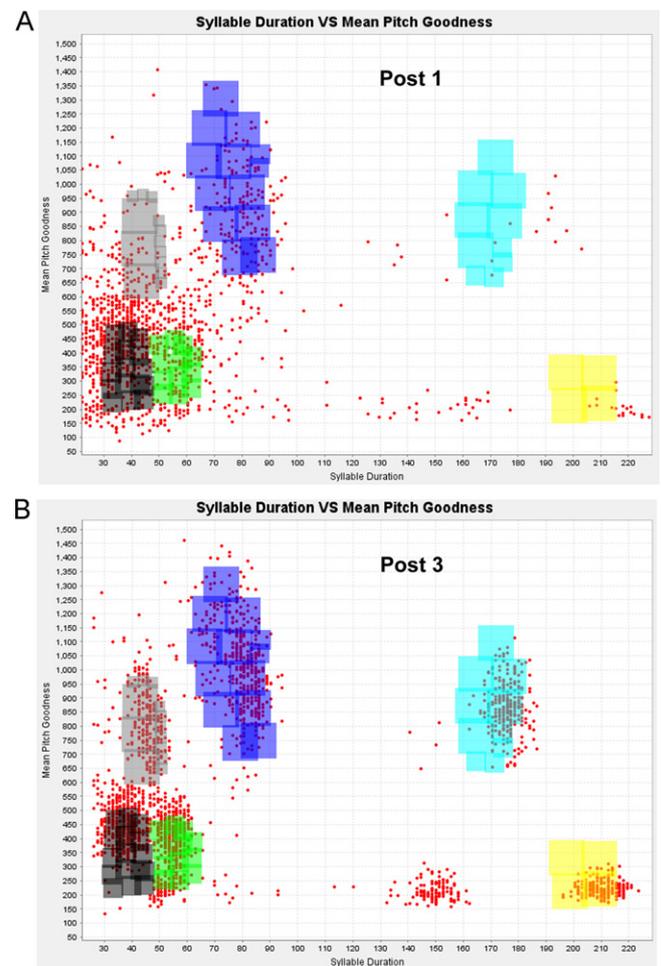


Fig. 5. The scatter plots of postoperative days 1 (A) and 3 (B) are superimposed with the painted clusters of the template. On the first day after a microlesion was made to the HVC, many notes were produced that did not fall within the syllable boundaries (A). On the third day after surgery, more notes were within the syllable boundaries (B). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

the number of different notes in the bout by the number of transition types in the bout. Since the number of different notes in a bout is always less than or equal to the number of transitions in the bout, a value of 1 represents the best (highest) linearity score. The consistency score is calculated by dividing the sum of typical transitions in the bout by the sum of the total transitions in the bout (a value of 1 represents the best consistency score). A stereotypy score is then calculated as the average of the linearity and consistency scores. Scores are calculated for every song bout and then averaged over all bouts. The scores are then written into a new spreadsheet.

2.4. Kullback–Leibler (K–L) distance analysis

The Kullback–Leibler distance (K–L distance) is a measure of the difference between two probability distribution functions (Wu et al., 2008). We use this to compare syllable transition probability distributions. That is, we calculate the K–L distance between the transition probabilities of two different days to quantify the dissimilarity of the syllable sequences on those days. For one example included in Section 3, we were interested in the disruption of the song, and song recovery, after partially lesioning the vocal motor pathway, so we compared the transition probability distributions of postoperative days of singing to that of the preoperative day (day 1). Let n denote the number of syllables along with NMS (e.g.,

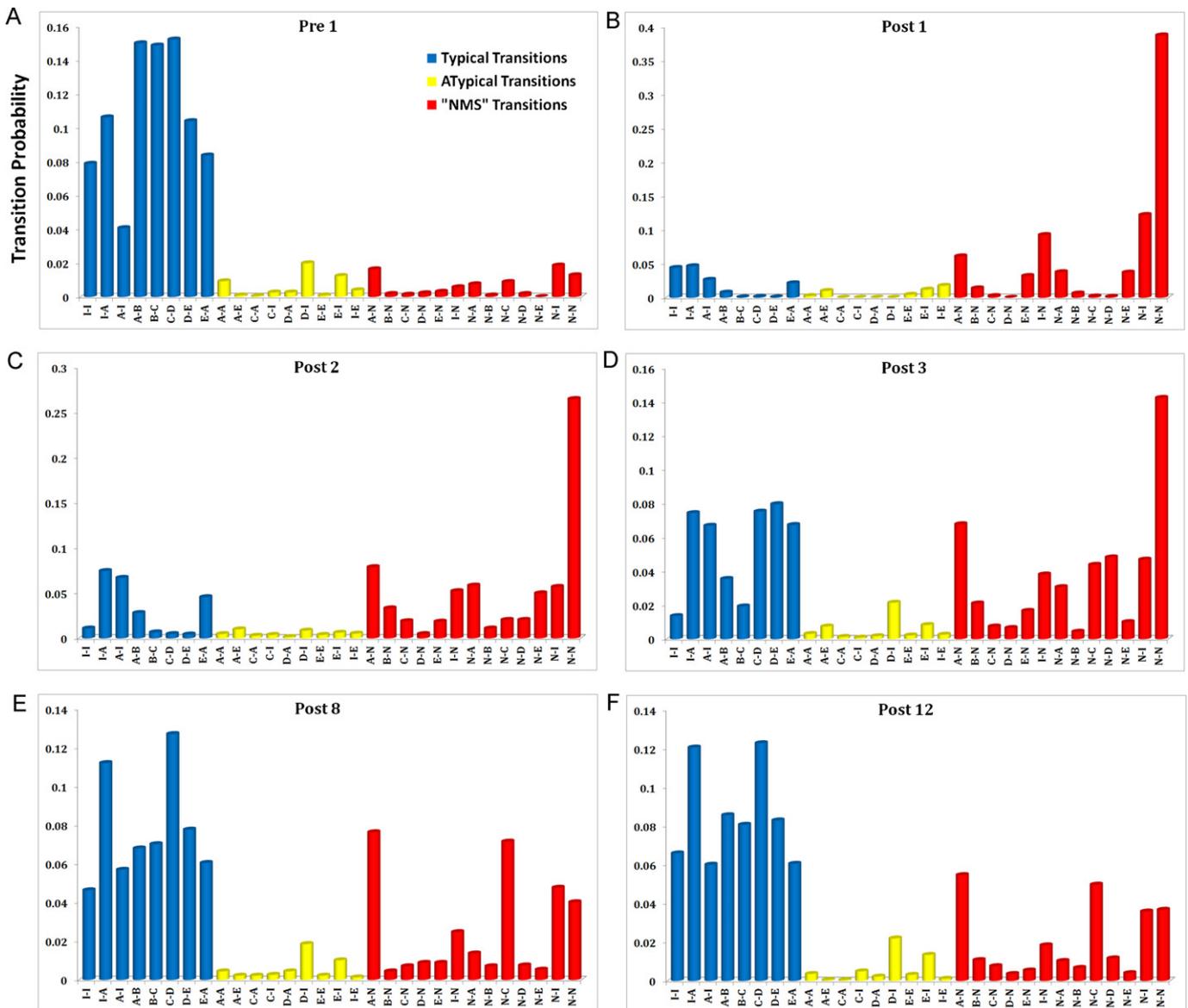


Fig. 6. Syllable transition probabilities are shown on preoperative day 1 (A), and on postoperative days 1 (B), 2 (C), 3 (D), 8 (E), and 12 (F). Blue bars represent “typical transitions”, yellow bars represent “atypical transitions”, or transitions that were not listed by the user as typical, and the red bars represent “NMS transitions” involving notes that lie outside the syllable boundaries. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

if the syllables entered by the user are A, B and C, then $n = 4$). There are $m = 2^n$ possible combinations of two-syllable transitions. If a transition never occurs, we set its probability to a non-significant small value (10^{-4}) to make the computation numerically stable and accurate. Let $T_1^1, T_2^1, \dots, T_m^1$ represent the transition probabilities on the preoperative day, and let $T_1^k, T_2^k, \dots, T_m^k$ represent the transition probabilities on day k . For example, if T_1^1 is the probability of an A to B transition on day 1, then T_1^3 is the probability of an A to B transition on day 3 (e.g., Fig. 6 shows the transition probabilities on one preoperative and five postoperative days). Then the K–L distance quantifying the degree of dissimilarity of the syllable sequences between days 1 and k is given by the following formula:

$$D_{K-L}^{1,k} = \sum_{j=1}^m T_j^1 \log_2 \left(\frac{T_j^1}{T_j^k} \right)$$

A K–L distance score is generated for every target file entered by comparing its transition probabilities with those of the first

target file. The scores are then plotted in ascending order of the sequence of target files entered. Fig. 7A shows a K–L distance analysis of the syllable sequence from one bird over two preoperative and 12 postoperative days of singing. The K–L distance of the first day of singing (Pre1) with itself is 0. Other data points describe the K–L distance between Pre1 and days following Pre1. This gives the time course of sequence dissimilarity over the days following Pre1. Larger values of the K–L distance reflect greater dissimilarity in the sequence.

2.5. Transition entropy analysis

The entropy of a probability distribution is a measure of the spread of that distribution (Wu et al., 2008). In our context, a high entropy means that there are many syllable transitions with non-negligible probabilities. A song has low entropy if only a few syllable transitions typically occur. If $T_1^k, T_2^k, \dots, T_m^k$ represent the transition

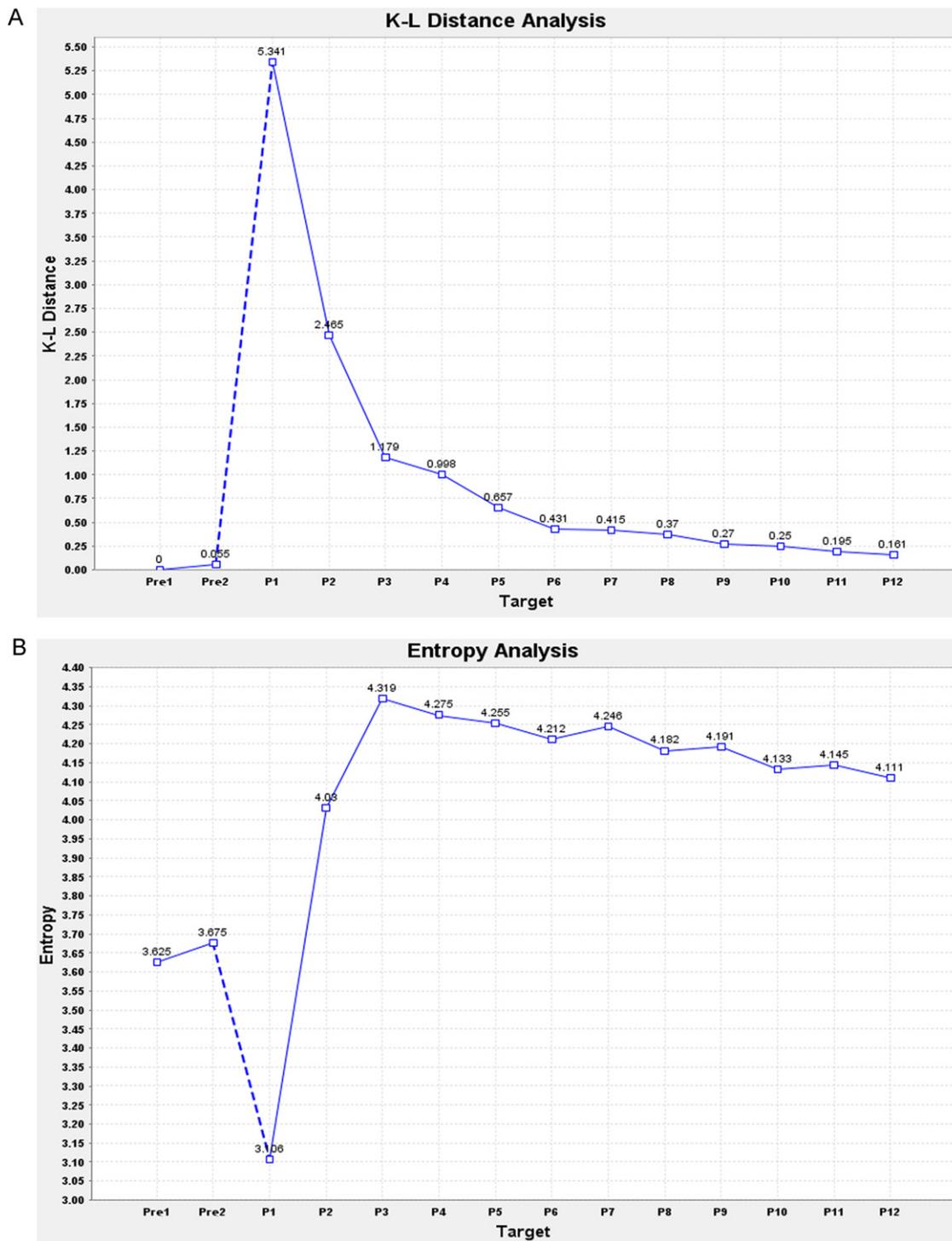


Fig. 7. Quantification of syllable transition distributions. (A) K-L distance measure quantifying the dissimilarity in song sequence between the Pre and Post days. Pre1 and Pre2 are preoperative days of singing, and P1–P12 are the days of postoperative singing. (B) Entropy analysis quantifying the spread of the syllable transition distributions for each day of singing. Dashed lines indicate the day of surgery.

probabilities on day k , then the entropy corresponding to that day is given by the formula:

$$D_k = \sum_{j=1}^m T_j^k \log_2(T_j^k)$$

The entropies corresponding to the transition probabilities of all target files are plotted in ascending order of the sequence of target files entered (Fig. 7B).

3. Results

To illustrate the results obtained by *SongSeq*, we use two behavioral data sets from different male zebra finches. The first bird, an adult, received bilateral microlesions targeted at the song region HVC (proper name). HVC contains projection neurons that contribute to the pathway that descends from the telencephalon to the brainstem vocal and respiratory network. This pathway is known as the vocal motor pathway (VMP) and is necessary for the

acquisition and the production of song. HVC neurons have a temporal role in song production via HVC-RA projections (Long et al., 2010 and reviewed in Bolhuis et al., 2010). The bird that we consider had a targeted microlesion that removed only a small region of the HVC (5–10%, Thompson et al., 2007), resulting in a disruption of his song post-surgery that gradually recovered to the preoperative state over a period of days. Next, we consider a second bird that had been recorded over the course of development. With the help of *SongSeq*, we monitor the bird's vocal changes and the evolution of his song over the course of development.

3.1. Quantifying the effects of HVC microlesions on syllable sequence

Fig. 2A shows a sonogram of a typical song for an adult bird displaying three introductory notes (labeled as I) and five syllables (labeled as A, B, C, D, and E) repeated over multiple motifs. Fig. 2B shows a 2D scatter plot of two selected acoustic features from a preoperative day of singing (syllable duration versus mean pitch goodness). The six note clusters correspond to the five syllables repeated in song motifs and some introductory notes (the colored regions in Fig. 2C). On the first day after a microlesion was made to the HVC, the song became very disorganized and many notes were produced that did not fall within the syllable boundaries (Fig. 5A). On the third day after surgery, the bird song started to recover and more notes were within the syllable boundaries (Fig. 5B).

Fig. 6A shows the syllable transition probabilities on the day before the surgery (the probability of every transition “X–Y” is calculated by dividing the number of occurrences for “X–Y” by the total number of transitions during that day of singing). The first eight blue bars are “typical” syllable transitions within a motif while the last blue bar (E–A) is the transition from the last syllable of a motif to the first syllable of the next motif. The yellow bars represent transitions that were not listed by the user as typical, based on an initial examination of the sonograms. Finally, the red bars

indicate transitions involving “NMS” notes that lie outside syllable boundaries.

During the first day of singing following the HVC microlesion, motif transitions occur with low probabilities and non-motif transitions occur more frequently (Fig. 6B). In fact, the highest transition probability is from one NMS note to another. During the second postoperative day of singing (Fig. 6C), the probabilities of motif transitions increased (particularly transitions “I–A”, “A–I”, “A–B” and “E–A”) and the NMS transitions became more scattered. The frequent “N–N” transition that occurred on Post1 became less frequent and more “N–X” and “X–N” transitions began to occur as more notes fell within the syllable boundaries. The number of “NMS” transitions continue to decline during the third and eighth postoperative days of singing (Fig. 6D and E) while the “I–A”, “A–I”, “A–B”, “B–C”, “C–D”, “D–E” and “E–A” transitions increased dramatically. By postoperative day 12 (Fig. 6F), the overall pattern of syllable transition probabilities appears similar to the preoperative structure, although there are more “NMS” transitions on Post12 than on Pre1. *SongSeq* can also show the transition distributions in terms of pie charts (not shown).

The dissimilarity in song sequence between the Pre and Post days is quantified using the K–L distance measure in Fig. 7A. Here, Pre1 and Pre2 are preoperative days of singing, and P1–P12 are the days of postoperative singing. Preoperative transition distribution functions are highly similar, so the K–L distance between Pre1 and Pre2 is near 0. On the first day of singing following surgery (P1) the K–L distance increases dramatically, since the typical transitions that occur on Pre1 are infrequent on P1. There is a big drop in the K–L distance between P1 and P2 due to recovery of the “I–A”, “A–B”, “A–I” and “E–A” transitions (Fig. 6B and C). Another big drop occurs between P2 and P3 due to the recovery of the “B–C”, “C–D” and “D–E” transitions and the further recovery of the “I–A”, “A–B”, “A–I” and “E–A” transitions (Fig. 6C and D). There is a third, but smaller, drop between P3 and P8 (Fig. 6D and E), making the typical transition probabilities in P8 more like those in Pre1. On subsequent days there is little change in the K–L distance.

Post 12

Average Linearity Score = 0.5981
Average Consistency Score = 0.6791
Average Stereotypy Score = 0.6386

.Wav File	Number of Transitions	Number of Transition Types	Linearity Score	Consistency Score	Stereotypy Score
00062.wav	30	13	0.5385	0.8	0.6692
00063.wav	23	13	0.5385	0.6957	0.6171
00064.wav	22	17	0.4118	0.5909	0.5013
00065.wav	18	13	0.5385	0.7222	0.6303
00066.wav	18	11	0.5455	0.5	0.5227
00067.wav	20	12	0.5833	0.4	0.4917
00068.wav	24	15	0.4	0.2917	0.3458
00069.wav	18	12	0.5833	0.7222	0.6528
00070.wav	23	12	0.5	0.7391	0.6196
00071.wav	22	11	0.6364	0.8182	0.7273
00072.wav	18	12	0.5833	0.7778	0.6806
00073.wav	7	6	0.8333	1.0	0.9167
00074.wav	14	8	0.75	1.0	0.875
00075.wav	23	11	0.5455	0.3478	0.4466
00076.wav	9	8	0.875	0.5556	0.7153
00077.wav	21	9	0.7778	0.9048	0.8413
00078.wav	19	10	0.6	0.6842	0.6421
00079.wav	18	10	0.7	0.7778	0.7389
00080.wav	18	11	0.6364	0.7778	0.7071
00081.wav	16	9	0.6667	0.9375	0.8021
00082.wav	16	14	0.5	0.4375	0.4688
00083.wav	12	11	0.5455	0.5833	0.5644
00084.wav	24	10	0.7	0.7917	0.7458
00085.wav	12	8	0.75	0.5833	0.6667
00086.wav	18	11	0.6364	0.7222	0.6793
00087.wav	20	12	0.5833	0.7	0.6417
00088.wav	26	12	0.5	0.8462	0.6731
00089.wav	14	10	0.7	0.8571	0.7786
00090.wav	27	17	0.4118	0.3333	0.3725

Fig. 8. Linearity, consistency and stereotypy scores for postoperative day 12. For every .wav file in the spreadsheet of day 12, the number of transitions as well as the number of transition types is calculated. Corresponding linearity, consistency and stereotypy scores are then generated and listed in tabular form. For any score that is larger than 0.85, the entry in the table is colored in red, indicating that the song for this .wav file is highly stereotyped. Average linearity, consistency and stereotypy scores are then calculated and shown at the top. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Fig. 7B shows a quantification of the spread of the transition probability distributions for each day of singing using the distribution entropy. The transition entropy values for Pre1 and Pre2 are similar, showing that the spread of the transition distribution function is similar on these two days. That is, the number of syllables and the order in which they are sung are similar. The entropy exhibited a dramatic drop on day P1 because the transition distribution on P1 has much less spread than that on any of the Pre days. Many notes sung on P1 were classified as NMS, so that many syllable transitions were lost. In fact, the NMS–NMS transition was most frequent on P1 (Fig. 6B). By day P2 many syllables have returned, but they are coupled with “NMS” notes, resulting in a transition distribution with a larger spread as quantified by the larger entropy. The entropy on day P3 increases further as more typical transitions are recovered and the “NMS” transitions declined (Fig. 6C and D). The entropy on day P3 is higher than that on the Pre days since, in addition to the typical transitions, there are more “NMS” transitions on day P3 than on the Pre days. The entropy declines slightly after that and reaches a plateau that remains somewhat above the Pre days. This example illustrates the different types of information encoded in the K–L distance and entropy measures. Moreover, by examining the scatter plots for target files used to generate the entropy and the 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 the template clusters), or some combination of the two.

Fig. 8 shows linearity, consistency, and stereotypy scores (Scharff and Nottebohm, 1991) generated for postoperative day 12. For every .wav file in the spreadsheet of day 12, the number of transitions is calculated as well as the number of transition types (the number of different transitions). Corresponding linearity, consistency and stereotypy scores are then generated based on these values. For any score that is larger than 0.85, the entry in the table is colored red, indicating that the song for this .wav file is highly stereotyped. Average linearity, consistency and stereotypy scores are then calculated and shown at the top. This procedure is done for

every target file entered and thus for every preoperative and postoperative day entered by the user. Fig. 9 shows a comparison of the average scores over all preoperative and postoperative days of singing (numbers removed for clarity). The microlesion clearly had a large impact on these measures of the song sequence during P1 and P2. The average consistency scores were higher than the average linearity scores on the Pre days. However, after surgery, and for the first five post operative days, the average linearity scores exhibit higher values than the average consistency scores. On day P6 the scores are almost equal, and for postoperative days six till twelve the consistency scores are again higher. As the bird song recovers the average scores increase gradually toward those on Pre 1.

3.2. Developmental changes in syllable sequence

In this next example we show the changes in the sequence of syllables sung by a bird during development. Fig. 10A shows the 2D scatter plot of two selected acoustic features (syllable duration versus mean FM) for a day of singing from an adult male zebra finch. The data points on the scatter plot form five clusters (Fig. 10B). There are five different motif syllables for this bird, and the canonical motif for this adult bird is ABCDCE. Thus, the typical transitions are “A–B”, “B–C”, “C–D”, “D–C”, “C–E” and “E–A”. The bird also often sings AABCDCE, making “A–A” an additional typical transition of his song. During the first week of development, the song is very disorganized and most of the notes produced do not fall within the syllable boundaries (Fig. 10C). However, during the subsequent weeks of development, the bird song started to shape up gradually and merge toward the adult song, and by the sixth week most notes were within the syllable boundaries (Fig. 10D).

The transition probabilities for the developing bird are shown in Fig. 11. As an adult, the bird song has high probability of motif syllable transitions and a low probability for non-motif syllable transitions (Fig. 11A). In contrast, in the same bird as a juvenile, during a day of the first week of singing (post-hatch days 35–42) there are few motif transitions and many transitions associated

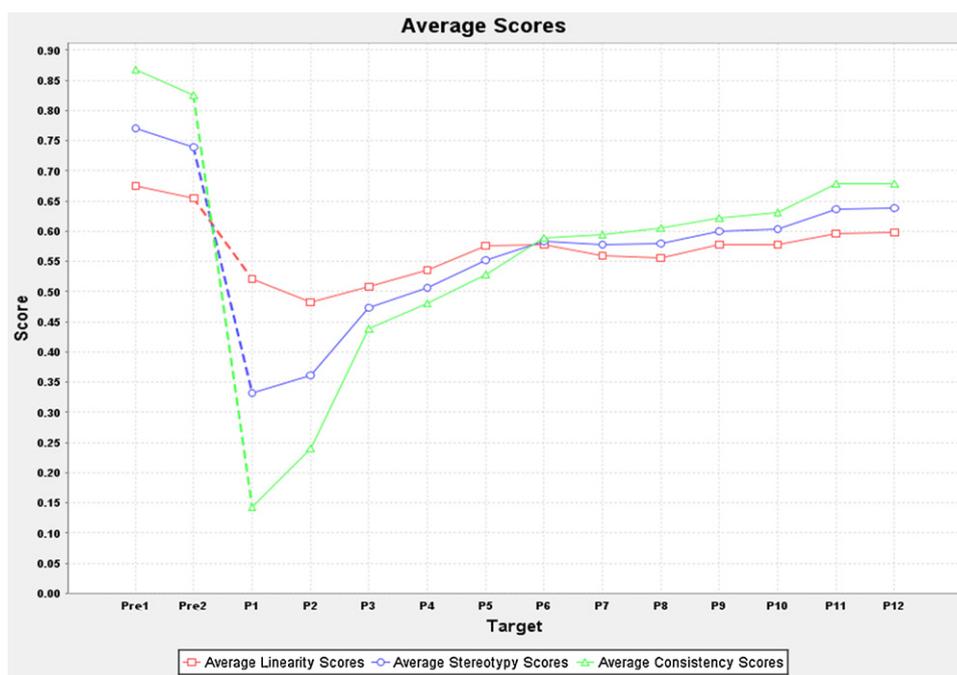


Fig. 9. Comparison of the average scores over all preoperative and postoperative days of singing. The microlesion clearly had a large impact on these measures of the song sequence during P1 and P2. However, as the bird song recovers the average scores increase gradually and by P12 they have returned to values similar to those on Pre 1. Numbers above data points are removed for clarity.

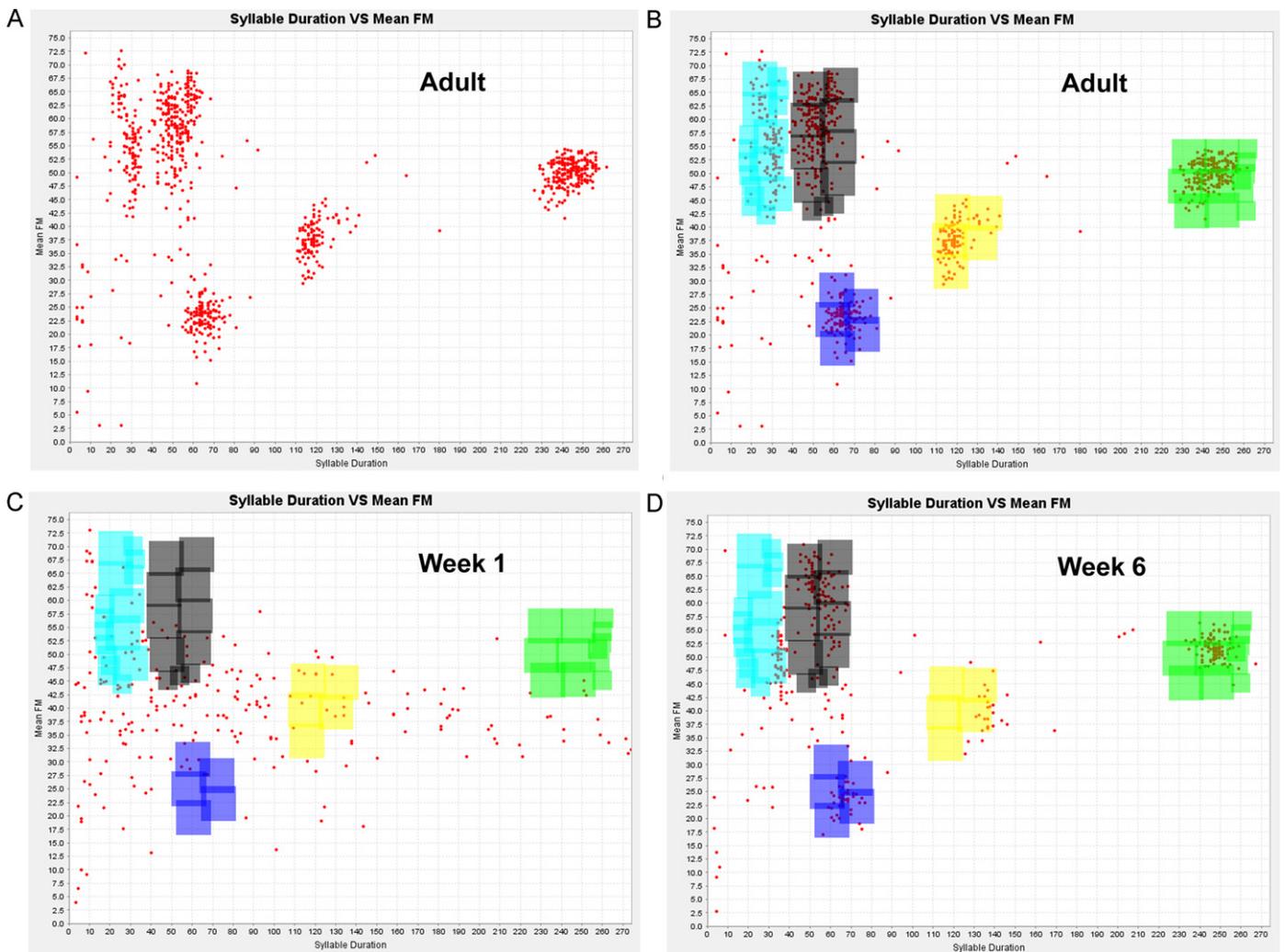


Fig. 10. Syllable identification for a developing bird. (A) The scatter plot of two selected acoustic features (syllable duration versus mean FM) for a day of singing from the bird when adult. (B) The data points on the scatter plot form five clusters indicating five different motif syllables for this bird that form the template. The scatter plots for weeks 1 (C) and 6 (D) are superimposed with the painted clusters of the template. During the first week of development, many notes were produced that did not fall within the syllable boundaries (C). During the sixth week, more notes were within the syllable boundaries (D).

with notes that fall outside the boundaries of the adult syllable clusters (Fig. 11B). The overall pattern of syllable transitions during weeks 2 and 3 was similar to week 1, exhibiting many transitions between an adult syllable and a “NMS” note (Fig. 11C and D). At a day during week 6 of singing in the juvenile (Fig. 11E), the probabilities of motif transitions increased and the “NMS” transitions became less frequent. By week 9 (Fig. 11F), the overall pattern of syllable transitions appears similar to the adult syntax structure: high probability for motif syllable transitions and low probability for non-motif transitions.

Fig. 12A shows the vocal change in the syllable sequence over the course of development of the bird as quantified by K–L distance analysis. This is done by comparing transition probabilities of the juvenile bird from days of different weeks of singing with that of the adult bird. That is, the K–L distance at target W_i compares the transition distribution from one day of week i with that of the adult (the template). On a day during the first week of singing (W1), the K–L distance is large since few of the adult typical transitions are made by the juvenile bird. This is true also for singing on W2 and W3; almost all transitions involve “NMS” notes. Therefore, the K–L distance remains elevated, and is even larger on W3 than W2 due to a decline in the “B–C” transition from W2 to W3 (Fig. 11C and D). By W6, many of the adult transitions are present (Fig. 11E), so the K–L distance is much lower.

Fig. 12B shows the transition entropy analysis quantifying the spread of the transition probability distributions for each week of singing. The entropy value on W1 is smaller compared with the adult, as one can observe by the smaller distribution spread in Fig. 11B versus Fig. 11A. On subsequent weeks of singing, the entropy values first rise as more syllables emerge in the song, and then decline as the transition sequence becomes more like that of the adult. By W9 the entropy is approximately the same as that of the adult bird.

Average linearity, consistency and stereotypy scores for the juvenile versus the adult songs are shown in Fig. 13. The linearity score on W1 is almost the same as that of the adult, yet the song is very different. The high linearity on W1 occurs because the bird sings many single or double notes. The difference between the adult and W1 songs becomes evident, however, when one considers the consistency score, which is almost 0 on week 1. By W9 both linearity and consistency scores are similar to those of the adult song.

3.3. Robustness

The results shown in our examples are robust to variations in the number of notes. To demonstrate this, we applied *SongSeq* on both birds’ data sets after removing 45% of the rows in each of the

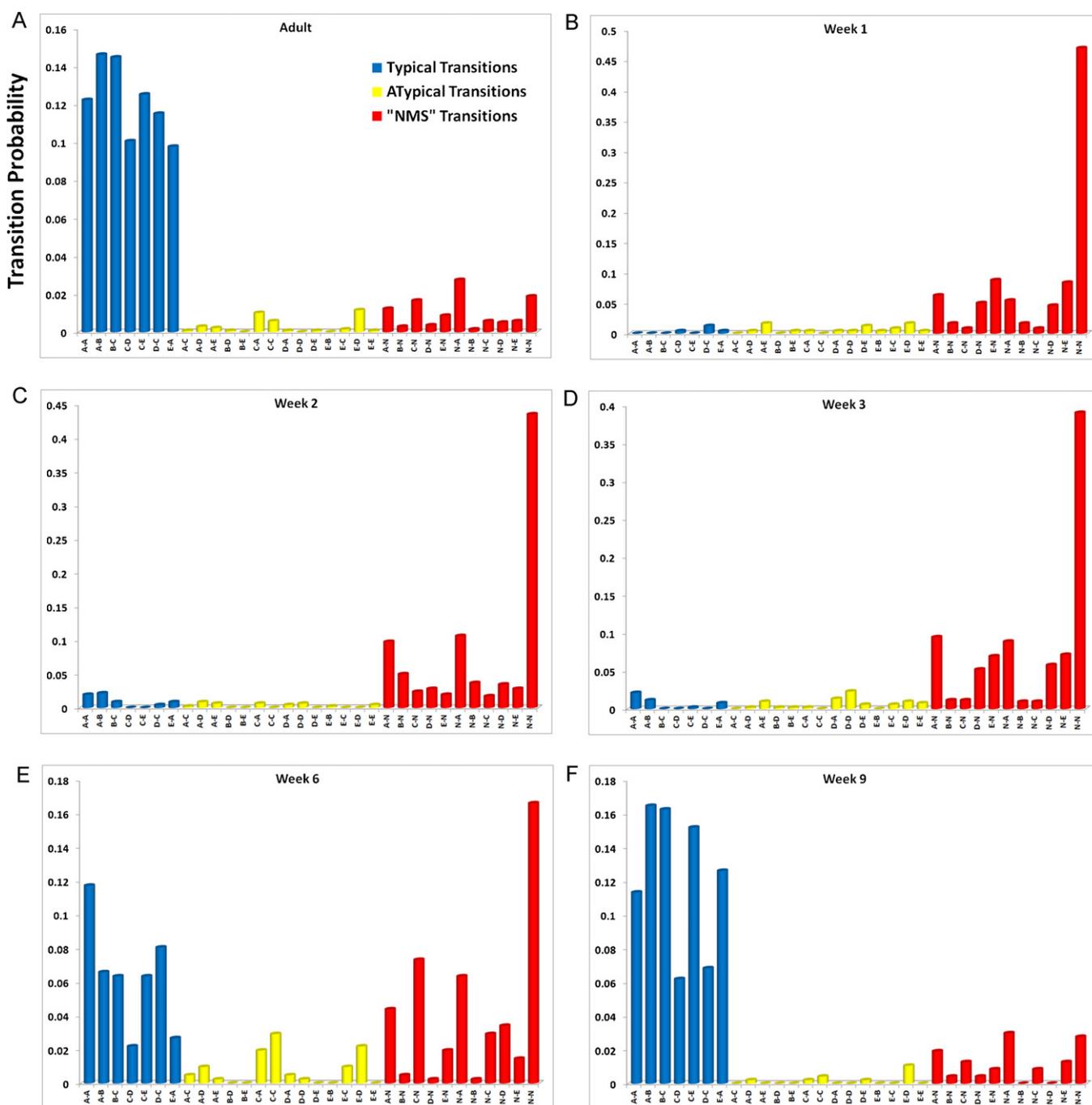


Fig. 11. Syllable transition probabilities are shown during a day when the bird is an adult (A), and during a day of week 1 (B), 2 (C), 3 (D), 6 (E), and 9 (F).

spreadsheets generated by Sound Analysis Pro (which removed 40–50% of the bouts on each day). The painted regions (and thus the syllable boundaries) on the template's 2D scatter plot remained the same using a module of *SongSeq* that allows the user to upload a previous template (*SongSeq* saves the dimensions of every painted box on the template's scatter plot each time a simulation is run). Fig. 14 shows the K–L distance analysis (Fig. 14A), entropy analysis (Fig. 14B) and average scores (Fig. 14C) for the first bird obtained after removing roughly half of the song bouts. The time courses of these measures are quite similar to those from the same bird when all song bouts were included (Figs. 7 and 9). Comparisons were also favorable when the same procedure was applied to the bird used in the development study (Fig. 15, compare with Figs. 12 and 13).

4. Discussion

We have presented an automated tool, *SongSeq*, for analyzing birdsong syllable sequences. We showed how *SongSeq* can be used to monitor changes of sound features across a large number of songs, analyze transition distributions among syllables, quantify syllable ordering in terms of linearity, consistency and stereotypy scores, and quantify the degree of similarity in song syntax over different days of singing. The software uses two standard measurements from information theory, the Kullback–Leibler distance and entropy, to quantify the transition distributions and compare day-to-day differences.

The Similarity Batch module in Sound Analysis Pro is a commonly used method to measure bird song similarity. For example,

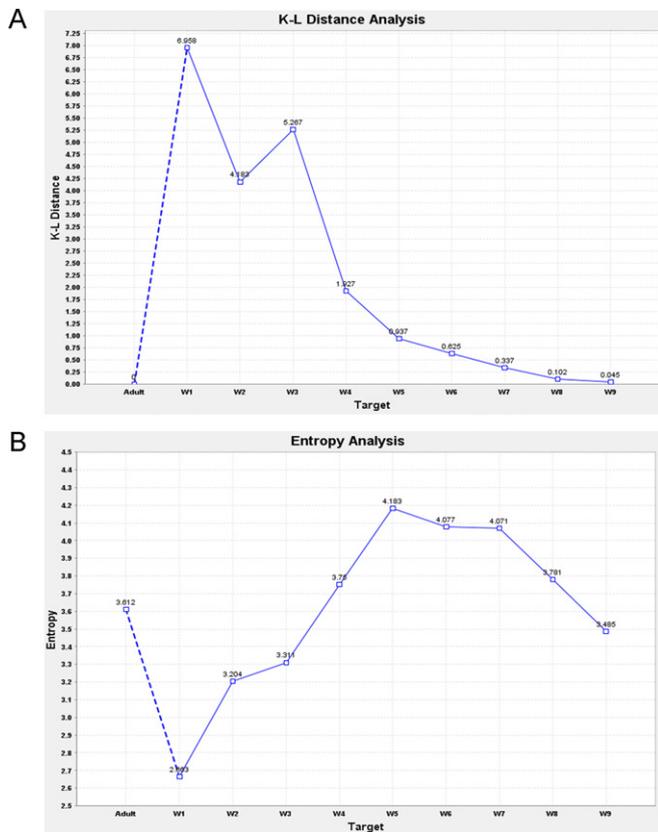


Fig. 12. (A) During weeks 1–3, the K–L distance is large since the juvenile song is quite different from the adult song. By week 4, however, the sequence was much improved. (B) The entropy value on W1 is smaller compared with adult (Fig. 9B versus Fig. 9A). On subsequent weeks of singing, the entropy values first rise as more syllables emerge in the song, and then decline as the transition sequence becomes more like that of the adult.

the method has been used to assess the vocal imitation of pupils (bird learning the song) from tutors (adult birds) (Tchernichovski et al., 2000) or recovery of song following brain injury (Coleman and Vu, 2005; Thompson et al., 2007). Similarity Batch can be used to perform a large set of similarity measurements. It supports two batch modes: one is for comparing ordered pairs of sounds, and the other is for comparing sound matrices. This module 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, *SongSeq*’s algorithm is based on the individual syllable transitions; the entropy and K–L distance functions are based on the distributions of acoustic features of individual syllable transitions, and thus, dissimilarity between two transition probability distributions can be traced back to individual syllable contributions.

SongSeq also provides an easy automated way to generate the linearity, consistency, and stereotypy scores. To the best of our knowledge, the only automated tool available to compute stereotypy scores is the web-based program located at (<http://bottjerlab.usc.edu/songinator.html>). This tool computes linearity and consistency statistics for bird song but requires the user to manually enter the syllable order for every song file, and thus it requires user inspection of every sonogram. In contrast, *SongSeq* allows large scale computation of the stereotypy scores over a large number of .wav files (and thereby song motifs) and over a large number of days of singing (and thereby multiple SA+ spreadsheets). This is done without user interaction and the only requirement from the user is to paint the clusters in the input stage to identify the syllables. Moreover, the average scores computed at the end of the batch process (Figs. 9 and 13) provides an informative view of the behavior of the bird’s song over multiple days of singing.

There is accumulating evidence that the bird song is coded at the sub-syllabic level (Day et al., 2009; Ravbar et al., 2012). During development, some subsyllables could form sooner than others, so



Fig. 13. Average linearity, consistency and stereotypy scores for the juvenile versus the adult songs. Here, W1–W9 represents a day during each of weeks 1–9.

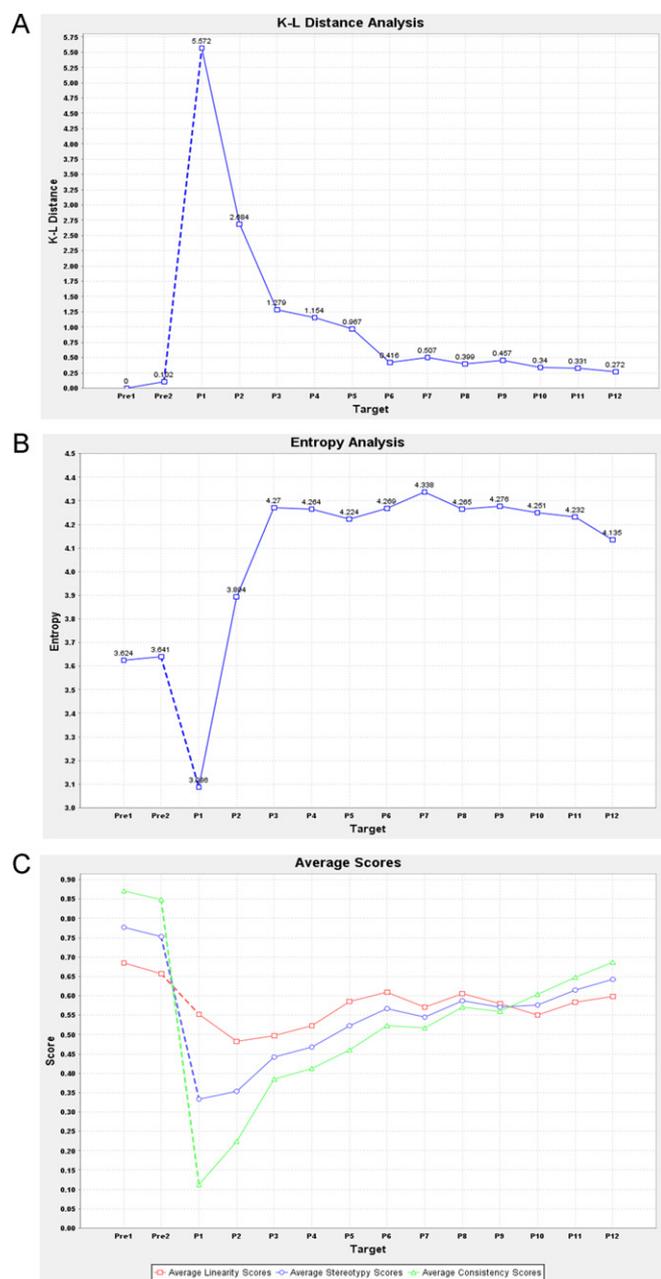


Fig. 14. Robustness of the results to variations in the number of notes. *SongSeq* was used on the bird 1 (HVC microlesion) data sets after removing 45% of the rows in each of the spreadsheets generated by Sound Analysis Pro. The K–L distance analysis (A), entropy analysis (B) and average scores (C) obtained after removing 45% of the song bouts. The time courses of these measures are quite similar to those from the same bird when all song bouts were included (Figs. 7 and 9).

that a syllable X' could form early that is close to, but outside, the boundaries of adult syllable X. Rather than counting these notes in the “NMS” category, one could define the variant X' in the template, by simply painting in the region of the template scatter plot where X' is needed. Then in the adult song, there would be X syllables but few X' syllables, while during development there would be many variants X' of X recorded.

The *SongSeq* software makes the identification of syllables much less labor intensive. However, there will always be misidentification of notes, just as there would be with manual identification. To reduce the frequency of misidentification, it is best to keep the syllable boundaries relatively tight. This is facilitated by using small painting boxes. A cloud of notes near, but not contiguous with, a

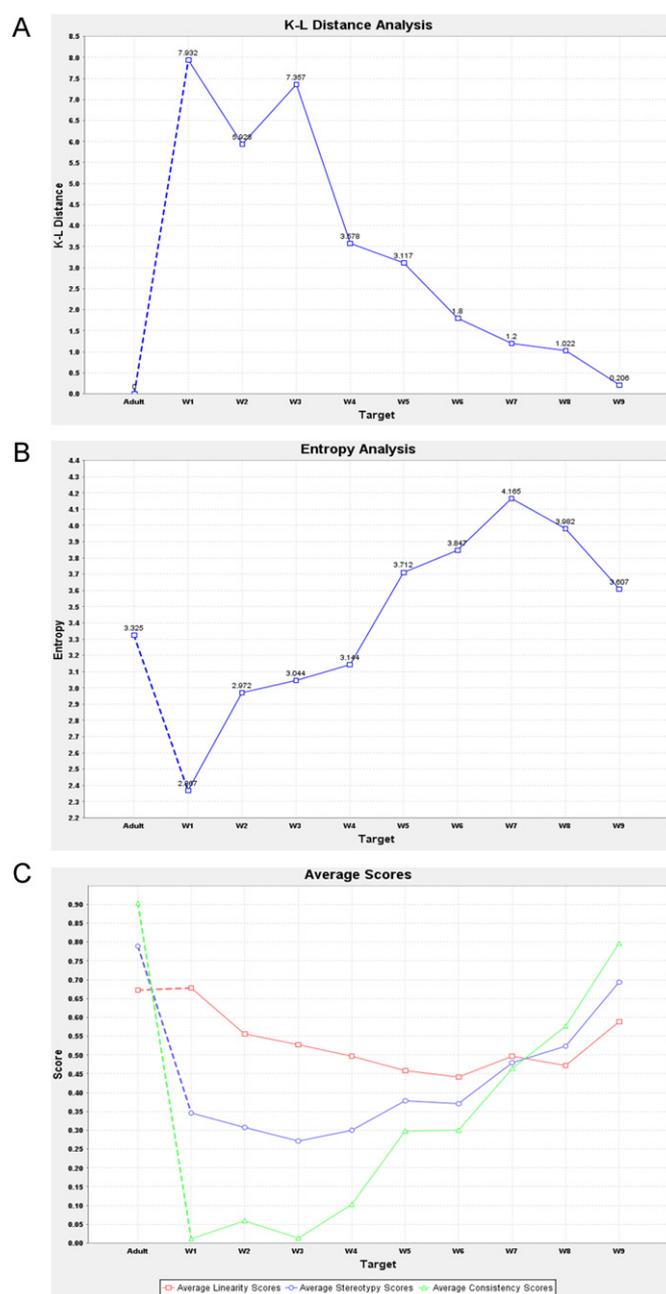


Fig. 15. Robustness of the results to variations in the number of notes. *SongSeq* was used on the developing bird data sets. The K–L distance analysis (A), entropy analysis (B) and average scores (C) obtained after removing 45% of the song bouts. The time courses of these measures are quite similar to those from the same bird when all song bouts were included (Figs. 12 and 13).

note cluster may best be categorized as a variant X' of the main syllable cluster X, as described above, rather than extending the boundaries of X. Also, by combining several days of singing one can get a clearer view of the syllable clusters, since the number of notes in the combined scatter plot is greater. This assumes, of course, that the song is similar on each day of singing that is combined (for example, combining days of singing from an adult bird prior to surgery).

In summary, *SongSeq* automates the most labor intensive components of bird song analysis. In addition to automated syllable identification, it employs several algorithms to quantify song syntax. The user interface is through a GUI that requires no user programming or data manipulation. The software, along with

online tutorial and a user manual, are available for free download at <http://www.math.fsu.edu/~bertram/software/songbird>.

Acknowledgement

This work was supported by grant number IOS-1146607 from the National Science Foundation.

References

- Bolhuis JJ, Okanoya K, Scharff C. Twitter evolution: converging mechanisms in bird-song and human speech. *Nat Rev Neurosci* 2010;11:747–59.
- 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.
- Day NF, Kinnischtzke AK, Adam M, Nick TA. Daily and developmental modulation of “premotor” activity in the birdsong system. *Dev Neurobiol* 2009;69:796–810.
- Long MA, Jin DZ, Fee MS. Support for a synaptic chain model of neuronal sequence generation. *Nature* 2010;468:394–9.
- Ravbar P, Lipkind D, Parra LC, Tchernichovski O. Vocal exploration is locally regulated during song learning. *J Neurosci* 2012;32:3422–32.
- Scharff C, Nottebohm F. A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *J Neurosci* 1991;11:2896–913.
- 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.
- 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.
- Wu W, Thompson JA, Bertram R, Johnson F. A statistical method for quantifying songbird phonology and syntax. *J Neurosci Methods* 2008;174:147–54.