Title:

A quantitative method for analyzing species-specific vocal sequence pattern and its developmental dynamics

Raimu Imai[1], Azusa Sawai[1], Shin Hayase[1], Hiroyuki Furukawa[1], Chinweike Norman Asogwa[1], Miguel Sanchez[1], Hongdi Wang[1], Chihiro Mori[1], and Kazuhiro Wada[1, 2, 3 #]

[1]Graduate School of Life Science, [2]Department of Biological Sciences, and [3]Faculty of Science, Hokkaido University, Sapporo, Hokkaido, Japan, 060-810

Corresponding author:

Kazuhiro Wada, MD. PhD.

Associate Professor,

Faculty of Science, Department of Biological Sciences,

Hokkaido University

Room 910, Building No.5, North 10, West 8, Kita-ku

Sapporo, Hokkaido 060-0810, Japan

E-mail: wada@sci.hokudai.ac.jp

Office Phone/Fax#: +81-11-706-4443

Conflict of interest: There are no conflicts of interest.

Abstract

Background: Songbirds are a preeminent animal model for understanding the neural basis underlying the development and evolution of a complex learned behavior, bird song. However, only a few quantitative methods exist to analyze these species-specific sequential behaviors in multiple species using the same calculation method.

New method: We report a method of analysis that focuses on calculating the frequency of characteristic syllable transitions in songs. This method comprises two steps: The first step involves forming correlation matrices of syllable similarity scores, named syllable similarity matrices (SSMs); these are obtained by calculating the round-robin comparison of all the syllables in two songs, while maintaining the sequential order of syllables in the songs. In the second step, each occurrence rate of three patterns of binarized “2 rows × 2 columns” cells in the SSMs is calculated to extract information on the characteristic syllable transitions.

Results: The SSM analysis method allowed obtaining species-specific features of song patterns and intraspecies individual variability simultaneously. Furthermore, it enabled quantitative tracking of the developmental trajectory of the syllable sequence patterns.

Comparison with existing method: This method enables us to extract the species-specific song patterns and dissect the regulation of song syntax development without human-biased procedures for syllable identification. This method can be adapted to study the acoustic communication systems in several animal species, such as insects and mammals.

Conclusions: This present method provides a comprehensive qualitative approach for understanding the regulation of species specificity and its development in vocal learning.

Keywords

vocal learning; vocal development; zebra finch; syntax; song learning; species specificity; individual variation.

Highlights

- New method is developed for discrimination of species specificity of vocalization.
- Prevalence of the characteristic syllable transitions is calculated.
- It quantitatively evaluates inter- and intraspecific differences of bird song.
- The method allows tracking the developmental trajectories of song patterns.

1. Introduction

The acquisition of species-specific behavioral patterns and their subsequent lifelong maintenance is a crucial research theme in order to understand the diversity and evolution of complex learned behaviors. Specifically, species-specific behaviors, such as mating, foraging, nest construction, and social communication, are generally composed of a rendition of temporal sequence patterns of behavioral units. These sequential behavioral characteristics are represented with species specificity even between closely related species (Berridge, 1990; Donaldson and Young, 2008; Greenspan and Ferveur, 2000; Weber and Hoekstra, 2009; Weber et al., 2013). Therefore, the sequential behavioral pattern is a critical species-specific biosignal for the recognition of conspecifics and the discrimination of heterospecifics in nature (Leininger and Kelley, 2015; Marler and Peters, 1988; Pollack and Hoy, 1979; Woolley and Moore, 2011). However, a limited number of methods have been developed for quantitative analysis of the temporal sequence of behavioral patterns.

Over 3,000 species of songbirds around the world have species-specific song patterns. These patterns comprise stereotyped acoustic units, called syllables, which are arranged in characteristic species-specific sequential orders. Similar to that of human language, songbirds learn such song features through auditory and vocal experiences in a defined sensitive period during which juvenile songbirds listen, memorize, and gradually match their own developing vocalizations to the song pattern of an adult tutor (Konishi, 1965; Tchernichovski et al., 2001; Waser and Marler, 1977). An understanding of how syllables are sequentially arranged to form songs patterns is of general interest. This is because the learning and ordering of stereotyped motor patterns is the basis of vocal communication in vertebrates including speech in human. They thus may underlie other aspects of motor control. However, there exist some limitations on the analysis of vocal phenotypes. This was

especially evident in the regulation of the temporal sequential patterns of syllables in many songbird species because of the difficulty in identifying each syllable (Gardner et al., 2005; Liu and Nottebohm, 2007). Acoustic features of syllables are variable in multiple factors, such as duration, frequency modulation, amplitude, and pitch, across song renditions and even in a song bout (Tchernichovski et al., 2000). Nevertheless, for the analysis of sequence or syntax of the song, categorization of syllable types was often performed based on human visual inspection of the spectrogram shapes. Although a few studies adapted semi-automatic clustering of syllables based on their mean acoustic features (Deregnaucourt et al., 2004; Ravbar et al., 2012; Wu et al., 2008), it remains difficult to set clear standards for identifying unique song syllables. This misleads the researcher given the observer-dependent bias on the identification of syllables, and precludes the extraction of precise information on the temporal sequence pattern of syllables in the song. A similar problem also occurs in the analysis of vocal patterns in mammals. For example the ultrasonic vocalizations produced by many species of rodents consist to a great extent of acoustically variable syllables (Holy and Guo, 2005). In addition, during vocal development, highly variable acoustic features of syllables are observed at the subsong and plastic song stages before the period of vocal crystallization. This makes it difficult to extract the quantitative information on the development of the temporal sequence of syllables.

Here, we adapted a correlation matrix capturing information on the similarity between syllables in two songs, termed as “syllable similarity matrix (SSM),” to extract songs’ temporal structures by quantification of specific syllable transition types rather than transitions of identified syllables. This method allows the detection of the species specificity of song temporal patterns, and quantitative tracking of the developmental trajectories without human-biased procedures in syllable identification.

2. Materials and Methods

2.1. Animals

All zebra finches (ZF, *Taeniopygia guttata*) and the Bengalese finches (BF, *Lonchura striata* var. *domestica*) were laboratory bred. Other species, such as owl finch (OF, *Taeniopygia bichenovii*), star finch (SF, *Neochmia ruficauda*), Java sparrow (JS, *Padda oryzivora*), and canary (CN, *Serinus canaria*) were obtained from local breeders. The photoperiod was constantly maintained at a 13:11 h light/dark cycle with food and water provided *ad libitum*. All bird experiments were performed according to the guidelines specified by the Committee on Animal Experiments at Hokkaido University. The guidelines are based on the national regulations for animal welfare in Japan (Law for the Humane Treatment and Management of Animals; after partial amendment No. 68, 2005).

2.2 *Song recording and tutoring*

Song recordings were performed as previously reported (Mori and Wada, 2015). Briefly, the birds were individually housed in a sound-attenuation box. A microphone (SHURE microphone SM57, SHURE incorporated, IL, USA) was connected to a FirePod FP10 amplifier (PreSonus Audio Electronics Inc. Florida, USA). Songs were recorded at a sampling rate of 44 kHz and 16 bit amplitude resolution by using Sound Analysis Pro version 1.04 (Tchernichovski et al., 2001). Songs were then automatically saved for 24 h per day by using the Sound Analysis Pro program. Low-frequency and high-frequency noises (<0.5 kHz and >15.8 kHz, respectively) were filtered using Avisoft SASLab pro software (Avisoft Bioacoustics, Berlin Germany). Audacity's noise-canceling procedure (version 2.0.5) was used to further filter ventilator and fluorescent-delivered noises.

With respect to the song tutoring experiment, the sex of ZF and BF chicks was determined

by polymerase chain reaction within 5 post-hatching day (phd), as described previously (Wada et al., 2006). Between 6 and 15 phd, before juveniles could start memorizing a tutor song, they were separated from their fathers and were kept with their mothers and siblings until fledging. Song tutoring was started at 15–20 phd and continued through over 100 phd. After fledging, birds were individually housed in a sound-attenuating box attached containing a mirror to reduce social isolation. Tutor songs were played five times in the morning and five times in the afternoon at 55–75 decibels from a speaker (SRS-M30, SONY) passively controlled by Sound Analysis Pro (v1.04) (Tchernichovski et al., 2000).

2.3. Calculation of similarity score between syllables

In this study, a song bout was defined as a continuous production of songs separated by a silent period longer than four folds of the mean value of more than 30 randomly selected inter-syllable gaps. Syllables in the song bouts were used for making SSMs. More than 50 syllables in song bouts were used as a song rendition. Introductory notes in a song were not included in a song rendition. The songs of the ZF, OF, SF, BF, and JS usually contained less than 50 syllables in a song bout. In these cases, two or more song bouts were merged as a combined song rendition to contain more than 50 syllables. For the CN, songs with more than 200 syllables were used because their songs consisted of a considerable number of syllable repetitions when compared with that of the other species. Song spectrograms were formed as WAVE format files (.wav) by Avisoft SASLab Pro software. Each syllable in the song bouts was identified and individually saved as a WAVE format file by automated section labeling (threshold 0.001 V, hold time 0.01 s, and margin 0 s) by using Avisoft SASLab Pro software. Then, the machine-based separated syllables were double-checked by human eyes for the precision of syllable separation. The individual syllables were next saved

as their separated SON format files (.son) labeled with a unique name including information on its sequential ordering position in the song. The series of separated syllable files of two song renditions were transferred to the Avisoft CORRELATOR program for calculating the similarity scores between the syllables of two songs by the round-robin comparison. The score of syllable similarity was calculated as the peak correlation coefficient between two syllables. The two spectrograms of the separated syllables were shifted incrementally past each other along the time axis. For each offset position, the correlation coefficient is computed according to the following formula:

$$\Phi_{XY} = \frac{\sum_X \sum_Y ((a_{xy} - m_a) * (b_{xy} - m_b))}{\sqrt{\sum_X \sum_Y (a_{xy} - m_a)^2 * \sum_X \sum_Y (b_{xy} - m_b)^2}}$$

where m_a and m_b are the mean values of the spectrograms *a* and *b*, respectively. a_{xy} and b_{xy} are the intensities of the spectrogram points at the locations *x* and *y*, respectively. Syllable similarity score is a value ranging from 0 to 1. A value of 1 means that the two spectrograms are identical. A value of 0 means that there was no similarity between the spectrograms. **Supplementary material 1** is shown as an example comparison of similarity scores between syllables with variability in duration and other acoustic features taken from a plastic song sung by a juvenile bird.

Similarity scores between the syllables in two song renditions were exported into cells in Microsoft Excel spreadsheet by keeping the syllable sequence order in the original songs [1st step (left) in **Fig. 1A**]. The spreadsheet was named with the information of the similarity scores between the syllables as a SSM. In this study, 10 SSMs per bird were prepared by the round-robin comparison of five song renditions. To qualitatively visualize the information of syllable temporal sequences in songs, each cell in the SSM was color-coded according to the

value of the similarity score [1st step (right) in **Fig. 1A**].

2.4. Extraction of frequency of characteristic syllable transitions of interest.

For the quantitative analysis of syllable temporal structures, the occurrence rate of characteristic patterns of binarized "2 rows × 2 columns" cells in the SSMs was calculated. For the binarized patterning of 2 x 2 cells in the SSMs, the R software program was used to find the most similar binarized pattern for each 2 x 2 cell in the SSM from 12 possible patterns by comparison with the mean values 0.86 and 0.33 as similar (that was represented by black colors) and different (that was represented by white colors) syllables, respectively (R Core Team, 2013) (**Supplementary material 2**). The two mean values were defined as the most frequently observed values with two peaks in a total of 240,000 syllable similarity scores from six songbird species (ZF, OF, SF, BF, JS, and CN; n = 4 birds/each species, 1000 syllables/bird) (**Supplementary material 3**). As an alternative way for the binarization of cells in the SSMs, 0.55–0.60 of the similarity scores could be used as a threshold to distinguish similar (black) or different (white) syllables (**Supplementary material 4**). Then, for each SSM, the prevalence of each binarized 2 × 2 cell patterns in a SSM was calculated. In this study, particularly three specific types of 2 × 2 cells patterns were used as characteristic syllable transitions of interest (2nd step in **Fig. 1A**). Syllable transition type I was defined as a "paired syllables transition" indicating the existence of two successive syllables that were different but with same sequential order in two songs. This can be illustrated by "song 1 [· · A B· · · · ·] vs. song 2 [· · · · A B· · · ·]" (in this case, A and B represent two different syllables). Syllable transition type II was a case of existence of the "repetitive syllables transition" by two successive identical or very similar syllables in two songs. For instance "song 1 [· · · · · A A· ·] vs. song 2 [· · · A A· · · · ·]". Syllable transition type III was set as a

“nonmatching syllables transition” in the case where two successive syllables were not identical within a song and across two songs. An example of this is “song 1 [· · · · A B· · · ·] vs. song 2 [· · · C D· · · · ·]” (in this case, A, B, C, and D represented different syllables). The suitability of calculating each occurrence rate of three binarized 2 × 2 cells patterns as a syllable transition analysis was examined using artificial models of ZF, BF, and CN song patterns. The detection of species specificity of the model songs (**Fig. 1B**) was then calculated. The mean of the occurrence rate of the syllable transition types I, II, and III and their coefficients of variation (CV) from 10 total SSMs as an individual animal were used for statistical analyses.

3. Results

3.1 Visualization and extraction of species-specific temporal features of syllable sequence by the SSM analysis

Six species of songbird, namely zebra finch (ZF), owl finch (OF), star finch (SF), Bengalese finch (BF), Java sparrow (JS), and canary (CN) were selected to examine the potential of the SSM analysis method for extracting the species-specific features of the syllable temporal ordering from songs (**Fig. 2**). The six songbird species were represented by species-specific song patterns including motif structure, combination of syllable chunks and repetition, and repetitive phrases. For instance, SF songs represented the "motif-like" structure, which was not as strictly consistent across song renditions as the motif of ZF songs. The OF and JS produced repetitive syllables as a part of their songs. This was similar to the BF and CN. However, the numbers of repeated syllables and the number of repetitive syllables types were different across these species. Based on the observation of the species-specific features of songs in these species, the SSMs of the songs of each species were visualized as color-coded heat maps to observe the species-specific song temporal patterns qualitatively. The color-coded SSMs from each species represented distinct species-specific unique patterns (**Fig. 2**). The patterns of SSMs from the songs of the ZF and SF showed striped patterns that were derived from their motif structure. In contrast, the SSMs from the OF, BF, JS, and CN indicated darkened squares with variable sizes, which belonged to syllable-repetitive phrases in their songs. Furthermore, parts of the SSM from the JS songs indicated checkered-flag patterns. This suggested that there were alternative two syllable repetitions, like ABABABA (**Fig. 2**). Then, the prevalence of the characteristic three types of syllable transitions was calculated from binarized SSMs. These syllable transitions were named as transition type I, II, and III. Syllable transition type I included the consistent

sequential production of two different syllables between two songs, as “paired syllables transition”. In contrast, transition type II referred to a repetition of closely resembling syllables between two songs, as “repetitive syllables transition”. Transition type III was defined as a type of syllable transition belonging to different syllable transitions, termed as “nonmatching syllables transition”. Similar to the scenario demonstrated using artificial song model patterns (**Fig. 1B**), even in the cases that used real songs from these species, the SSM pattern analyses succeeded in extracting species-specific occurrence rates of the characteristic three transition types. Then, different appearance ratios of the transition types were observed among the species (**Fig. 2)**. For instance, the ZF song typically showed a higher value in the transition type I compared with those of other species. In contrast, OF, JS, and CN songs had higher values in the transition type II when compared with the other species. However, the three species possessed different values in the transition type III. This species specificity of the prevalence of the syllable transition types was clearly shown in three- or two-dimensional maps (**Fig. 3A** and **B**).

Additionally, these dimensional maps represented intraspecies individual variability of the temporal patterns. For example, in the ZF song, the motif structure was observed as a typical species-specific feature. Nevertheless, the number of syllable types composed in the structure ranging with 2 to 5 syllables was different across individuals. Such individual differences of the motif structure were shown as different values of the transition type I. That is, the motif structure formed with three syllable types showed higher values of type I than the motif structure formed with four syllable types. In the group data of each species, the value and ratio of mean occurrence rates of the three transition types and its CV were represented differently across the six species. This indicated species-specific traits of the syllable transition patterns (**Fig. 3C**). For example, it is known that some ZFs produce repetitive call-

like syllable transitions at the end of their song motifs. Such variability in syllable repetition was detected as a high CV value of syllable transition type II in the ZF (**Fig. 3 C**). These results implied that the analysis of the SSM-based method could be suitable for discriminating the species-specific features of syllable sequence transition in the crystallized songs at an adult stage in multiple songbird species.

3.2. Detection of the development of the species specificity of syllable transitions in a song

To elucidate the development of species-specific song patterns in songbirds, the SSM analysis method was further adapted for monitoring the developmental trajectory of the syllable transition of songs through the various song acquisition stages. It was a challenge to analyze the development of species-specific song patterns across multiple species by using the same platform of the analysis algorithm. This was due to the difficulty in quantitatively calculating the amorphous vocal patterns called subsong and plastic song. These were observed as intermediate states until song crystallization. The analysis using the SSM method did not require identification or alphabetization of syllables in songs. Therefore, the quantitative development of species-specific song patterns was analyzed regardless of the existence of highly variable acoustic features of syllables during subsong and plastic song stages.

For this purpose, song files from ZFs and BFs were used. These were recorded through development under playback tutoring conditions with conspecific songs (each n = 4) (**Fig. 4**). The results of the SSM analyses for the development of the species specificity of song patterns revealed that the songs of both ZF and BF juveniles did not indicate any characteristic patterns in the SSMs at 40–50 phd. They indicated low occurrence rates of both

paired- (transition type I) and repetitive- (transition type II) syllable transitions but with high scores of transition type III. Then, at 60–70 phd, the differences of the species specificity of song patterns appeared in the song developmental trajectory in the two-dimensional maps of transitions I and II between ZF and BF (**Fig. 5**). Interestingly, each juvenile from the two species showed an individually unique trajectory of song development with distinct combination of drastic changes and slumber states of both transition types I and II during 10 day intervals until 100 phd. For instance, bird ZF3 showed the largest change of the mean occurrence rate in transition type I between 50 and 60 phd. The bird then slowly crystallized its song pattern by the iteration of minor changes (ZF3 in **Fig. 4C**). Furthermore, individual variability in song learning strategy was also observed among birds keeping the same species-specific developmental constraints (**Fig. 4C** and **D**). Although both BF1 and BF4 represented unique individual trajectories of song development similar to the tutor song, BF1 increased only in the repetition transition shown as an accumulation of the value of the transition type II. However, BF4 increased in both paired (transition type I) and repetition (transition type II) transitions (BF1 and BF4 in **Fig. 4D**). In the species group data, the song development trajectory of the ZFs and BFs revealed a gradual increase in the species-specific prevalence of syllable transitions I and II, respectively, over the course of development (p = 0.00423 and p = 3.88E-06, respectively, two-way ANOVA). In contrast, the prevalence of the syllable transition type III, “nonmatching syllables transition”, continuously decreased during song development in the ZFs and BFs (**Fig. 5**, upper panels**)**. However, there was a trend toward differences in the mean number of Type III transitions in ZF vs. BF starting from an early developmental stage, with more transitions observed in BF than in ZF (**Fig. 5**). Higher values of Type III transitions stems from songs with more variable and unique syllables. Therefore, this difference in Type III transitions suggests that BF juveniles start and continue to sing

with more unstable syllables than ZF juveniles do over the course of song development. The CV of syllable transition types I, II, and III did not show significant differences between the ZFs and BFs. This suggests that both species had similar fluctuation rates in the prevalence of the syllable transitions (**Fig. 5**, lower panels**)**. These results indicated the potential of the SSM analysis for studying the species-specific vocal development and monitoring individual unique traits in a variety of species.

4. Discussion

The SSM analysis method enabled the extraction of the species-specific features in syllable sequence transitions. It also allowed the examination of the developmental process of the features during song learning in songbirds. The SSM analysis of song patterns is based on obtaining and processing the prevalence of selected characteristic types of the syllable transitions (**Fig. 1A**). Irrespective of its simplicity, the analysis revealed the species-specific distribution of syllable transition types among the species and showed intraspecies individual variation (**Figs. 3** and **4**). The CV of the prevalence of characteristic transition types is also useful for detecting species-specific variability of each syllable transition types (**Fig. 3 C**). In addition, CV values of transition types indicates behavioral fluctuations occurring during the song development, as shown by higher CV values of the syllable transition I and II in the ZF and BF, respectively, at juvenile vs. adult stages (**Fig. 5**).

To gain a precise understanding of song development, two vocal parameters are crucial in analyzing song phenotype, namely how a juvenile develops “syllable acoustics” as the behavioral units and “syllable sequence” as the control of the behavioral unit ordering. Recent computer-aided sound recordings and analyses led to significant improvements in quantitative song analysis, especially with respect to the study of developmental changes of the syllable acoustic phenotypes (Deregnaucourt et al., 2005; Tchernichovski et al., 2001). Although a few studies have adapted semi-automatic clustering of syllables based on their mean acoustic features, studying the development of the temporal sequence of syllables in songs remains a significant challenge (Daou et al., 2012; Lipkind et al., 2013; Ravbar et al., 2012; Wu et al., 2008). Many recent analyses of the syllable sequence are based on the songs with identified/alphabetized syllables (Horita et al., 2008; Okanoya and Yamaguchi, 1997; Scharff and Nottebohm, 1991), as it has been difficult to extract features of the syllable

transitions with acoustically ambiguous syllables. Therefore, the analysis of syllable sequences was constrained and used a limited number of songbird species which produced consistently stable syllables across song renditions (Gardner et al., 2005; Liu et al., 2004; Liu and Nottebohm, 2007; Marler and Slabbekoorn, 2004; Rose et al., 2004). A similar problem was observed in the analysis of the development of syllable sequences in the song. This was because of a difficulty in identifying acoustically unstable syllables/notes during song learning.

Unlike the previous analysis methods based on syllable categorization, the SSM method quantifies the occurrence rate of characteristic types of the syllable transitions, not transitions of identified specific syllables. Therefore, the SSM method quantitatively detects features of the temporal sequence of songs, including species-specific differences and developmental dynamics. For instance, a majority of syllables at around 40–50 phd of ZF and BF did not have any clear transition rules and were identified as Type III transitions (**Figs. 4A and B**). In contrast, some of other syllables in the same song renditions were detected as mixed population of Type I and II transitions, meaning of the initiation of the species-specificity. Detection of the onset of the species-specificity of vocal patterns is a unique strength of SSM method. Therefore, as a new quantitative method for analyzing species-specific vocal sequence patterns and their developmental dynamics, the SSM method can complement existing methods already adapted for quantification of specific syllable transitions. However, the SSM-based analysis method still has room for improvements to enable a more precise and fine detection of species-specific features in at least two points, the calculation of syllable similarity, and the pattern recognition of the SSM. With respect to the calculation of syllable similarity, parametric methods such as pitch and frequency modulation of syllables could provide more precise similarity scores than the methods based on the comparison of the

image patterns of spectrograms. This could be especially applicable in a case where syllables with complicated spectrogram structures are compared (Tchernichovski et al., 2000). As for the pattern recognition of the SSM, the threshold setting for cell binarization could be arbitrarily chosen by researchers, depending on their research aims. Two different threshold settings were compared for cell binarization: “the mean value of 0.86 and 0.33 (as used in this study)” and “the similarity score 0.60 threshold (as boundary similar and different).” Consistency of the occurrence rates of the syllable transition types I, II, and III was observed using these two different thresholds for individuals of six species (**Supplementary material 4**). However, any specific threshold set for cell binarization is problematic for the subsong of juvenile birds and for adults in a species with variable syllable performance. To solve this potential problem, instead of using binarized 2 × 2 cells patterns for the fine detection of the species-specific features from the SSM pattern, machine-learning algorithms, such as supervised and/or anomaly learning methods, could be powerful tools for precise pattern recognition from SSM without cell binarization. Additionally, our SSM analysis did not include information on the inter-syllable silent gap. This is a crucial factor for regulating the song tempo (Saar and Mitra, 2008; Sasahara et al., 2015). However, even the simplest method of focusing on a limited number of transition types of syllables in the SSMs, allowed the detection of species-specific song features and vocal development by monitoring individual unique traits in songbird species.

Notably, the experiment with the song tutoring in the ZF and BF by using the SSM analyses revealed that all juveniles learned a certain degree of the tutored song features. This indicated a developmental drifting approach to the tutored songs in the two-dimensional plot with the syllable transition types I and II (**Fig. 4C** and **D**). However, the developmental trajectory of the prevalence of the syllable transitions was unique among individuals. This represented

individual variability in learning strategy during song development (Liu et al., 2004). Therefore, these results indicated that SSM-based analysis is a useful method for studying the developmental strategy of acquisition of the sequential behavior and its individual variability. Furthermore, the SSM analysis could be used to assess the strength of vocal learning in comparing the song of tutees and tutors. In this comparison, the SSM analysis simultaneously extracts similarity information related to both syllable transition (sequence) and acoustics (spectral) learning between the tutor's model and tutee's developing songs. In addition, the SSM method could be potentially adapted for evaluating the effects of experimental manipulation on the regulation of behavioral patterns. The changes in syllable transition regulation caused by deafening procedures and the developmental abnormality of specific syllable transitions in socially isolated birds were successfully detected to-date. It will be reported as a separate study in the near future. Also, the SSM-based analysis for sequential behaviors could be adapted to analyze other vocal behaviors that are species-specifically regulated, such as mating songs in insects and ultrasonic vocal communications in rodents, as well as the development of speech acquisition in human babies.

The extraction of occurrence rate of the characteristic syllable transitions from the SSM patterns represents an effective approach to evaluate species-specific song patterns and to track the developmental trajectory of the temporal regulation of syllable sequence, taking into account the intraspecies individual variability.

Figures and figure legends

Figure 1. Syllable similarity matrix (SSM) method for detection of syllable temporal transitions

(A) The SSM method consisted of two steps. First, a correlation matrix including syllable similarity scores was generated by the round-robin comparison of all syllable comparisons in two songs, keeping the sequential order of the syllables in the songs. Second, the occurrence rate of three patterns of binarized "2 rows × 2 columns" cells in the SSM was calculated as a percentage of the characteristic syllable transition types I, II, and III (see **Materials and Methods**).

(B) Test analyses of the SSM-based method were conducted by using artificial song models mimicking the songs of the zebra finch (ZF), the Bengalese finch (BF), and the canary (CN) that included motif, chunk, and repetition structures, respectively, as their typical syllable transitions. The occurrence rates of the syllable transition types I, II, and III in the three song models are represented as bar graphs.

Figure 2. Visualization and extraction of the species-specificity of song temporal patterns by the SSM method

Examples of song spectrograms, color-coding SSMs, and mean occurrence rates of the syllable transition types I, II, and III in the SSMs of six species of the songbird. Their phylogenetic relationship was represented as a dendritic three. Two arrow lines on the sides of the SSMs represent parts of songs shown as two spectrograms. Color-coding of the SSMs indicates the value of the syllable similarity scores (darker brown color means higher similarity score between compared syllables). Scale bars (in the images of birds) = 2cm.

Figure 3. Species-specific features of the syllable transitions represented by the SSM analyses

(A) Three-dimensional plot showing the prevalence of the three transition types (transition patterns I, II, and III) of syllables in the six species of the songbirds (red: ZF; n = 6, gray: SF; n = 5, green: OF; n = 6, blue: BF; n = 6, purple: JS; n= 5, and orange: CN; n = 6).

(B) Two-dimensional plots for the prevalence of the syllable transition types shown in **(A)**. Each dot represents mean ± S.D. of 10 SSMs in a bird.

(C) Mean of the occurrence rate of the three types of the syllable transitions and its CV in six songbird species. Bar graphs represent mean ± S.E.M.

Figure 4. Developmental dynamics of the species specificity of the song patterns in ZF and BF

(A), (B) An example of developmental changes in the song patterns, color-coding SSMs, and prevalence of the syllable transition types I and II in ZF **(A)** and BF **(B)**, which were tutored with a playback conspecific song.

(C), (D) Individual developmental trajectories of the song patterns represented with the syllable transition types I and II in ZFs **(C)** and BFs **(D)** (each n = 4). Open circles denoted songs of adult ZFs (red) and BFs (blue). The red and blue colored cross marks presented tutored ZF and BF songs, respectively.

Figure 5. Developmental dynamics of the species specificity of the song patterns in ZF and BF

Developmental changes of the mean of occurrence rate of the syllable transition types I, II, and III and its CV in the ZF and BF (each n = 4, mean of type I; species x phd $F_{1, 6} = 3.797$, * $p < 0.01$, mean of type II; species x phd $F_{1, 6} = 8.724$, ** $p < 0.0001$, two-way ANOVA). Error bars = S.E.M.

Acknowledgments

We thank Drs. W-C. Liu, S. Kojima, M. Soma, T. Matsushima, and M. Mizunami for providing critical comments. This work was supported by Grant-in-Aid for Japan Society for the Promotion of Science Fellows (to R.I. and C.M.), the Japanese Government MEXT scholarship (to C.N.A. and H.W.), and the Asahi Glass Foundation and Grants-in-Aid for Scientific Research 225640097 and 25290063 (to K.W.).

References

Berridge KC. Comparative fine structure of action: rules of form and sequence in the grooming patterns of six rodent species. Behaviour, 1990: 21-56.

Daou A, Johnson F, Wu W, Bertram R. A computational tool for automated large-scale analysis and measurement of bird-song syntax. Journal of neuroscience methods, 2012; 210: 147-60.

Deregnaucourt S, Mitra PP, Feher O, Maul KK, Lints TJ, Tchernichovski O. Song development: in search of the error-signal. Annals of the New York Academy of Sciences, 2004; 1016: 364-76.

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

Donaldson ZR, Young LJ. Oxytocin, vasopressin, and the neurogenetics of sociality. Science, 2008; 322: 900-4.

Gardner TJ, Naef F, Nottebohm F. Freedom and rules: the acquisition and reprogramming of a bird's learned song. Science, 2005; 308: 1046-9.

Greenspan RJ, Ferveur J-F. Courtship in Drosophila. Annual Review of Genetics, 2000; 34: 205-32.

Holy TE, Guo Z. Ultrasonic songs of male mice. PLoS biology, 2005; 3: e386.

Horita H, Wada K, Jarvis ED. Early onset of deafening-induced song deterioration and differential requirements of the pallial-basal ganglia vocal pathway. The European journal of neuroscience, 2008; 28: 2519-32.

Konishi M. The role of auditory feedback i the control of vocalization in the white-crowned saprrow. Zeitschrift fur Tierpsychologie, 1965; 22: 770-83.

Leininger EC, Kelley DB. Evolution of Courtship Songs in Xenopus : Vocal Pattern Generation and Sound Production. Cytogenet Genome Res, 2015; 145: 302-14.

Lipkind D, Marcus GF, Bemis DK, Sasahara K, Jacoby N, Takahasi M, Suzuki K, Feher O, Ravbar P, Okanoya K, Tchernichovski O. Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. Nature, 2013.

Liu WC, Gardner TJ, Nottebohm F. Juvenile zebra finches can use multiple strategies to learn the same song. Proceedings of the National Academy of Sciences of the United States of America, 2004; 101: 18177-82.

Liu WC, Nottebohm F. A learning program that ensures prompt and versatile vocal imitation. Proceedings of the National Academy of Sciences of the United States of America, 2007; 104: 20398-403.

Marler P, Peters S. The role of song phonology and syntax in vocal learning preferences in song sparrow, Melospiza melodia. Ethology, 1988; 77: 125-49.

Marler P, Slabbekoorn H. Nature's Music: The Science of Birdsong. Elsevier Academic Press, 2004.

Mori C, Wada K. Audition-Independent Vocal Crystallization Associated with Intrinsic Developmental Gene Expression Dynamics. Journal of Neuroscience, 2015; 35: 878-89.

Okanoya K, Yamaguchi A. Adult Bengalese finches (Lonchura striata var. domestica) require real-time auditory feedback to produce normal song syntax. Journal of neurobiology, 1997; 33: 343-56.

Pollack GS, Hoy RR. Temporal pattern as a cue for species-specific calling song recognition in crickets. Science, 1979; 204: 429-32.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. 2013.

Ravbar P, Lipkind D, Parra LC, Tchernichovski O. Vocal exploration is locally regulated during song learning. The Journal of neuroscience : the official journal of the Society for Neuroscience, 2012; 32: 3422-32.

Rose GJ, Goller F, Gritton HJ, Plamondon SL, Baugh AT, Cooper BG. Species-typical songs in white-crowned sparrows tutored with only phrase pairs. Nature, 2004; 432: 753-8.

Saar S, Mitra PP. A technique for characterizing the development of rhythms in bird song. PloS one, 2008; 3: e1461.

Sasahara K, Tchernichovski O, Takahasi M, Suzuki K, Okanoya K. A rhythm landscape approach to the developmental dynamics of birdsong. Journal of the Royal Society, Interface / the Royal Society, 2015; 12.

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. The Journal of neuroscience : the official journal of the Society for Neuroscience, 1991; 11: 2896-913.

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.

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

Wada K, Howard JT, McConnell P, Whitney O, Lints T, Rivas MV, Horita H, Patterson MA, White SA, Scharff C, Haesler S, Zhao S, Sakaguchi H, Hagiwara M, Shiraki T, Hirozane-Kishikawa T, Skene P, Hayashizaki Y, Carninci P, Jarvis ED. A molecular neuroethological approach for identifying and characterizing a cascade of behaviorally regulated genes. Proceedings of the National Academy of Sciences of the United States of America, 2006; 103: 15212-7.

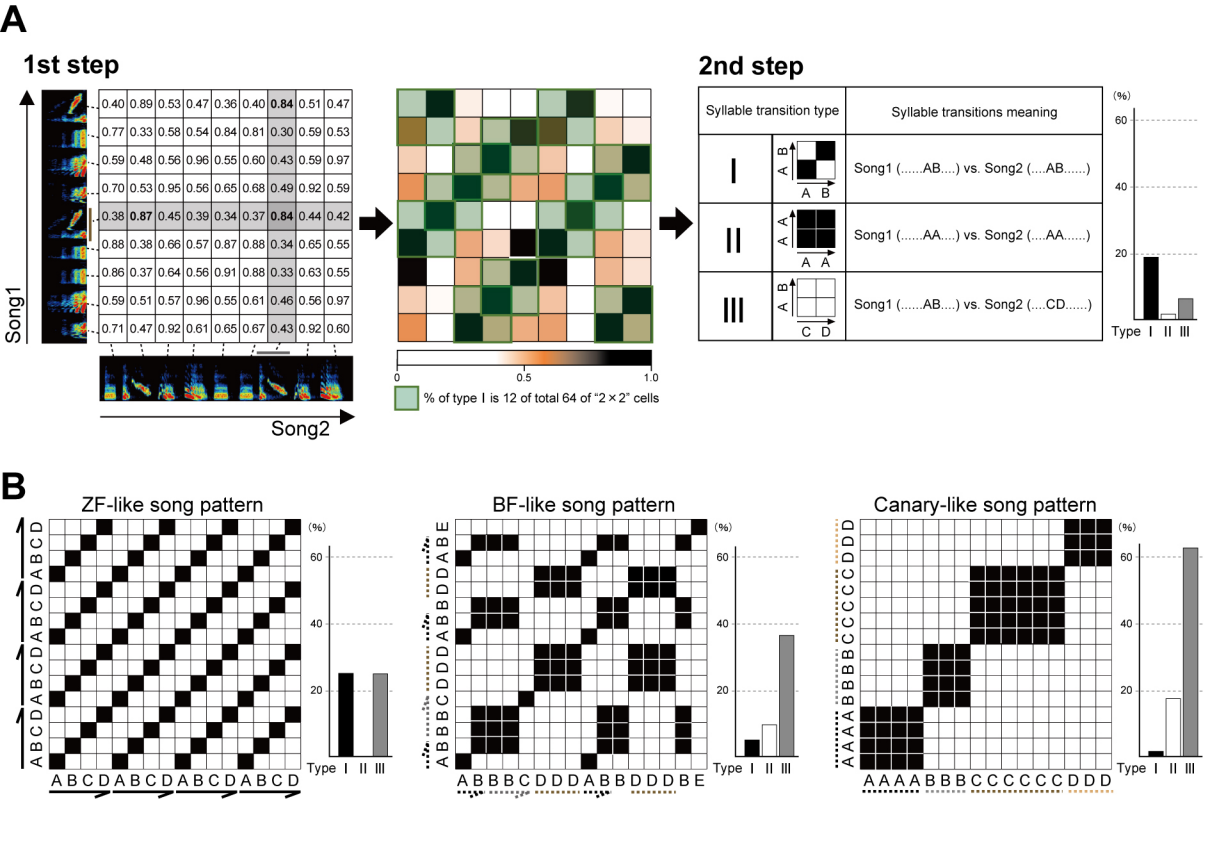
Waser M, Marler P. Song Learning in Canaries. JOURNAL OP COMPARATIVE AND PHYSIOLOGICAL PSYCHOLOGY, 1977; 91: 1-7.

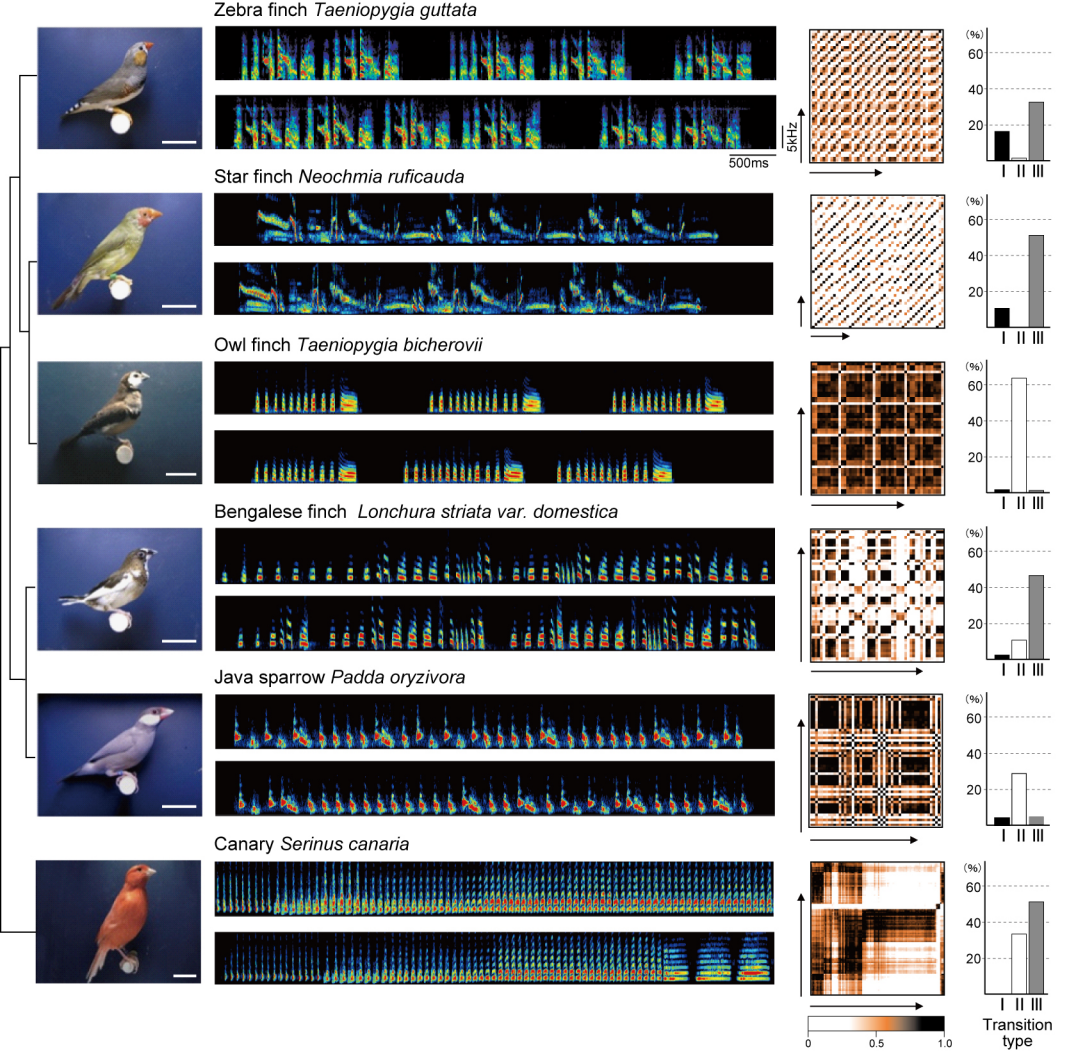
Weber JN, Hoekstra HE. The evolution of burrowing behaviour in deer mice (genus Peromyscus). Animal behaviour, 2009; 77: 603-9.

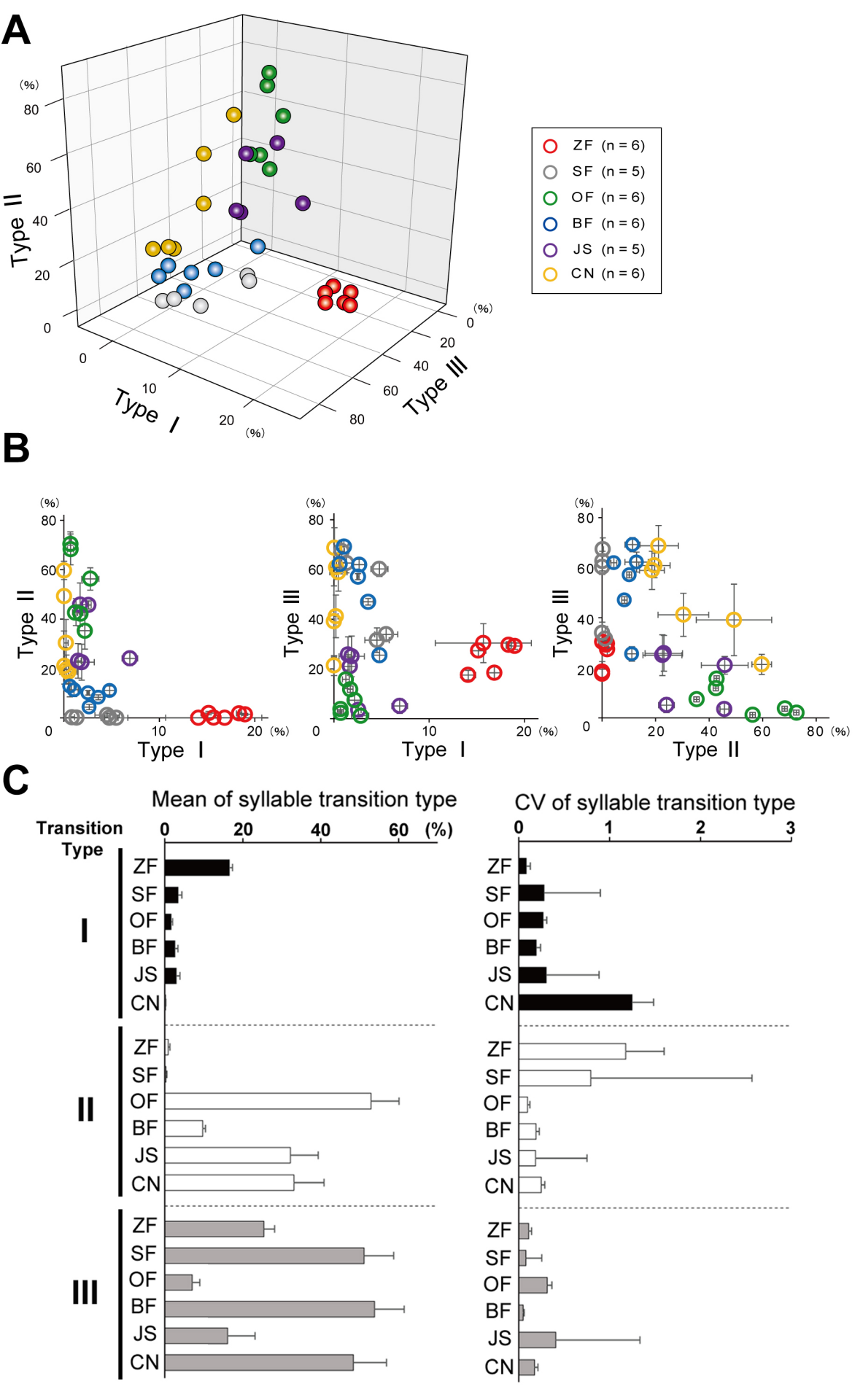
Weber JN, Peterson BK, Hoekstra HE. Discrete genetic modules are responsible for complex burrow evolution in Peromyscus mice. Nature, 2013; 493: 402-5.

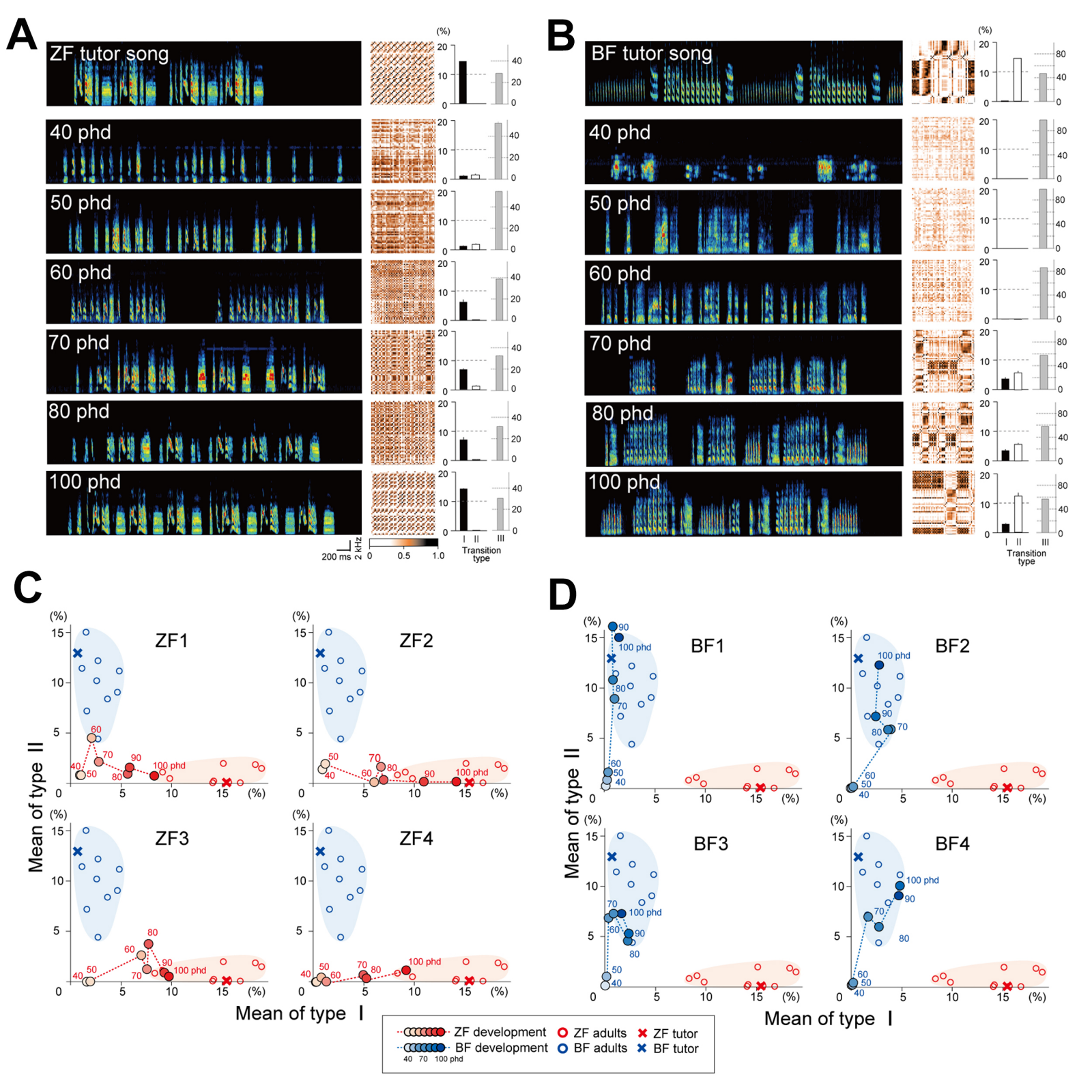
Woolley SM, Moore JM. Coevolution in communication senders and receivers: vocal behavior and auditory processing in multiple songbird species. Annals of the New York Academy of Sciences, 2011; 1225: 155-65.

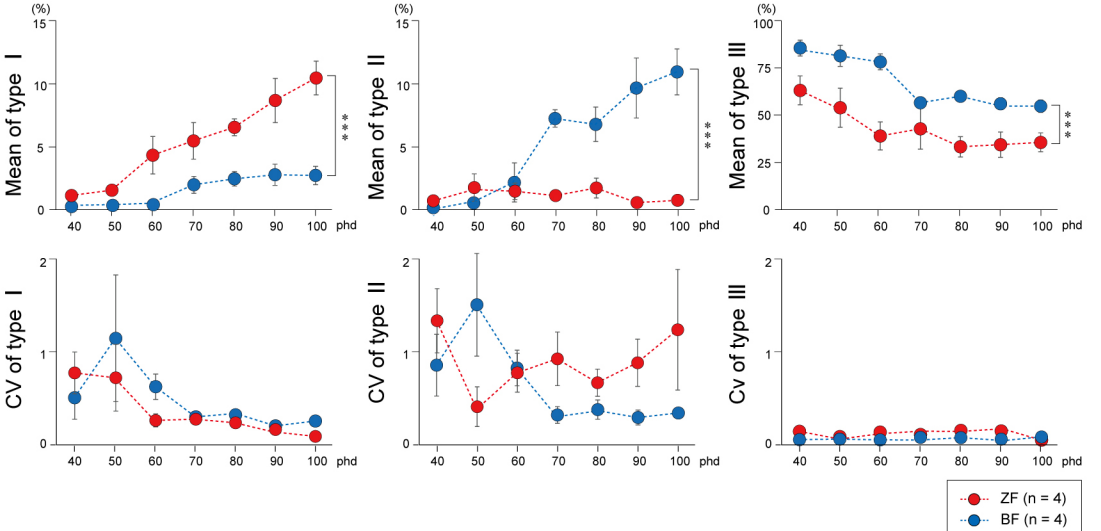
Wu W, Thompson JA, Bertram R, Johnson F. A statistical method for quantifying songbird phonology and syntax. Journal of neuroscience methods, 2008; 174: 147-54.


A
1st step
2nd step
Song1
Song2
0.40 0.89 0.53 0.47 0.36 0.40 0.84 0.51 0.47
0.77 0.33 0.58 0.54 0.84 0.81 0.30 0.59 0.53
0.59 0.48 0.56 0.96 0.55 0.60 0.43 0.59 0.97
0.70 0.53 0.95 0.56 0.65 0.68 0.49 0.92 0.59
0.38 0.87 0.45 0.39 0.34 0.37 0.84 0.44 0.42
0.88 0.38 0.66 0.57 0.87 0.88 0.34 0.65 0.55
0.86 0.37 0.64 0.56 0.91 0.88 0.33 0.63 0.55
0.59 0.51 0.57 0.96 0.55 0.61 0.46 0.56 0.97
0.71 0.47 0.92 0.61 0.65 0.67 0.43 0.92 0.60
0
0.5
1.0
% of type I is 12 of total 64 of "2 × 2" cells
Syllable transition type
Syllable transitions meaning
I
II
III
A B
A B
A A
A A
A B
C D
Song1 (.....AB....) vs. Song2 (....AB.....)
Song1 (.....AA....) vs. Song2 (....AA.....)
Song1 (.....AB....) vs. Song2 (....CD.....)
(%)
60
40
20
Type I II III
B
ZF-like song pattern
BF-like song pattern
Canary-like song pattern
A B C D A B C D A B C D A B C D
A B B B C D D D A B B D D D B E
A A A A B B B C C C C C C D D D
Type I II III


Zebra finch Taeniopygia guttata
5kHz
500ms
(%)
60
40
20
I II III
Star finch Neochmia ruficauda
Owl finch Taeniopygia bicherovii
Bengalese finch Lonchura striata var. domestica
Java sparrow Padda oryzivora
Canary Serinus canaria
0
0.5
1.0
Transition type


A
(%)
80
60
40
20
0
Type II
0
10
20
(%)
Type I
0
20
40
60
80
(%)
Type III
ZF (n = 6)
SF (n = 5)
OF (n = 6)
BF (n = 6)
JS (n = 5)
CN (n = 6)
B
Type II
Type I
Type III
Type I
Type III
Type II
C
Mean of syllable transition type
CV of syllable transition type
Transition Type
0
20
40
60
(%)
0
1
2
3
I
II
III
ZF
SF
OF
BF
JS
CN


A
ZF tutor song
40 phd
50 phd
60 phd
70 phd
80 phd
100 phd
(%)
200 ms
2 kHz
0
0.5
1.0
I
II
III
Transition type
B
BF tutor song
40 phd
50 phd
60 phd
70 phd
80 phd
100 phd
(%)
I
II
III
Transition type
C
ZF1
ZF2
ZF3
ZF4
Mean of type II
Mean of type I
(%)
D
BF1
BF2
BF3
BF4
Mean of type II
Mean of type I
(%)
ZF development
BF development
40 70 100 phd
ZF adults
BF adults
ZF tutor
BF tutor


(%)
Mean of type I
Mean of type II
Mean of type III
CV of type I
CV of type II
Cv of type III
***
40
50
60
70
80
90
100
phd
ZF (n = 4)
BF (n = 4)

Reviewers' comments: Review of Imai et al. (JNEUMETH-D-16-00159)

Reviewer #1— Summary:

The authors describe a novel statistical approach to analyze species differences and developmental changes to syllable sequencing. The method involves creating a spectral correlation matrix, followed by the discretization of neighboring syllables (as same or different), followed by a categorization of different types of syllable transitions. Thereafter, species differences or developmental changes in syllable patterning are visualized. The method produces interested and potentially powerful ways to study evolutionary and developmental variation in vocal communication patterns. The manuscript was overall well written and clear. Figures were generally useful.

Response:

We deeply appreciate the reviewer's careful reading of our manuscript and the constructive and supportive comments. Based on the reviewer's detailed suggestions, we revised the manuscript and figures as described below.

Reviewer #1— Major concern#1:

My first concern details with the generation of the spectral correlation matrix. The authors are thoughtful in extracting individual syllables from the songs then retaining the sequencing of these syllables in the correlation matrix. This allows one to eliminate variation in the timing of individual syllables in the computation of similarity between two songs. However, one complication is how to deal with variation in the duration of an individual syllable. For example, if a syllable is compare to a version of itself that is stretched by 10% (i.e., syllable is produced 10% slower), the spectral correlation would be reduced. It is not clear how much this could affect spectral correlation coefficients, but the authors should document the extent to which variation in the syllable durations could affect correlation coefficients and, ultimately, the pattern of sequence transition matrices. This will likely pose less of a problem for comparisons of adult songs than for the analysis of developmental changes in sequence patterning.

Response:

To address this concern, we added an example comparing syllable similarity scores between syllables with a variety of durations and those with other acoustic features. This example, derived, from a plastic song at juvenile stage, is shown as **Supplementary material 1** in our revised manuscript. Although the reviewer has expressed concern about a potential effect of variation in "syllable duration" on the spectral correlation coefficients, "syllable duration" is not the only factor to produce an effect on the similarity score between syllables. As shown in **Supplementary material 1**, when "syllable duration" differs between two syllables, other acoustic features, such

as mean FM and mean pitch, also usually differ. This is the reason of choosing to calculate syllable similarity based on the comparison of the image patterns of spectrograms in our study. As the reviewer pointed out, we also understand the potential limitations of the similarity calculation method in this study. Therefore, we added a discussion about this problematic point in the revised Discussion section (pages 16 and 17).

Reviewer #1— Major concern#2:

Another concern that I have pertains to the discretization of neighboring syllables as "similar" or "different". First, it would be useful to see examples of the distribution of correlation scores to determine the thresholds used. Second, it will be useful to examine how the results varied when the authors used the two different methods of defining "similar" vs. "different" (see Methods).

Response:

Based on the suggestions for the discretization of neighboring syllables, in **Supplementary material 3**, we now show the distribution of correlation scores of 240,000 syllables from six songbird species (n = 4 birds/each species, 1000 syllables/bird). These scores were used to determine the thresholds in this study.

To compare the results obtained with the use of two methods for defining “similar” vs. “different” syllables described in the Materials and Methods section, we re-calculated the occurrence rates of the syllable transition types I, II, and III of six species shown in Figure 2. On the binarization of cells in the SSMs with the threshold set at 0.60, the results indicated consistent values of the occurrence rates of the syllable transition types with those calculated by the threshold of mean values of 0.86 and 0.33 used in this study. The correlation between the results of two different threshold setting is shown as **Supplementary material 4**.

We added new sentences describing this analysis in our revised Materials and Methods and Discussion sections (pages 8 and 17).

Reviewer #1— Major concern#3:

Relatedly, I am wondering to what degree an analysis of the distribution of correlation scores might be useful (i.e., without discretization into similar vs. different). In other words, one could first plot the distribution of correlation scores between neighboring syllables and then describe the pattern of distributions across the species. In species with a high degree of syllable repetition, one would observe a distribution that was shifted to high correlation coefficients, whereas in other species with motif structure (e.g., zebra finch) one would find the distribution to be skewed toward lower correlation coefficients. One could then compute the K-L divergence (or other measures of difference) between the distribution of individuals within species to quantify intra-species variation, as well as differences between individuals of different species to quantify inter-specific

variation.

Response:

At the outset of our study, we had a similar idea for quantifying intra-species and inter-specific variations using the distribution pattern of syllable similarity scores to calculate K-L divergence. However, after several trials, we found problems associated with this analysis method. First, discretion for both intra-species and inter-specific variations was found to be less using this method than using our SSM method. In some cases, we observed that the values of intra-species variations were larger than those of inter-species variations. Second, this method excludes information related to syllable transitions. Therefore, it is difficult to extract information from species-specific song sequential structures using this approach. Because of these limitations, we did not adopt this analysis approach.

Reviewer #1— Major concern#4:

The authors should also discuss the results from their analysis of the CV of category types in more detail. In the current manuscript, analyses of CV are only sparsely mentioned, and the overall utility of these analyses remain unclear.

Response:

Based on this suggestion, we added text explaining the analysis of CV of the syllable transition types in our revised Discussion section (page 15).

Reviewer #1— Major concern#5:

The analyses of species differences in song development are interesting but more information is needed. First, more details on the nature of the tutoring is required. Were the song stimuli passively played back, or did juveniles actively control playback. (If the former, there seems to be a surprising degree of song learning in both species, at least in the examples provided in Figure 4). What are the thresholds to distinguish similar vs. different in this set of analyses? Were the same thresholds used as in the adult study?

Response:

For analyses of species differences in song development, the tutor songs were played back passively. We do not know whether our tutoring method results in a higher degree of song learning. However, to enhance juvenile song learning, we assessed the following factors during the song tutoring experiment: timings of the father's removal; initiation of song playback; total number of song playbacks in a day; and the placement of a mirror in a sound attenuation box to reduce social isolation. We added this information related to song tutoring methods to our revised Materials and Methods section (*2.2 Song recording and tutoring*). In this study, we used the same thresholds for all analyses, including those related to song development.

Reviewer #1— Major concern#6:

Second, the results from the developmental analysis need to be fleshed out. For example, it is interesting to note species differences in the mean number of Type III transitions from early on in development, but the authors do not discuss. Can these analyses be used to assess the strength of vocal learning (e.g., individual variation) within each species? To what degree does similarity in their measure of sequencing related to similarity in the learning of spectral features of syllables? This last analysis is not imperative for their paper but potentially of interest to distinguish sequence vs. spectral learning.

Response:

As the reviewer suggested, we have incorporated species-differences in the mean number of Type III transitions between ZF and BF in our revised Figure 5 (higher in BF than in ZF). This difference was observed from the developmental stage during subsong production. Higher Type III transitions stem from songs with more variable, less similar syllables. Therefore, this result suggests that BF juveniles start and continue to sing with more unstable and variable syllables than ZF juveniles. We added these findings in our revised Result section (pages 13 and 14).

We agree with the idea that the SSM analysis can be used to assess the strength of vocal learning in comparing the song of tutees and tutors. In this comparison, the SSM analysis extracts similarity information related to both syllable transition (sequence) and acoustics (spectral) learning, simultaneously. We have added this point to our revised Discussion section (page 18).

Reviewer #1— Minor comment#1:

I suggest the authors use "prevalence" to describe the frequency of transitions (e.g., in the "Highlights" section). "Frequency" is often used to describe the spectral structure of individual syllables, so it can be confusing for readers.

Response:

Based on this suggestion, we have replaced frequency (i.e., “appearance frequency of syllable transitions”) with either “prevalence” or “occurrence rate” in our revised manuscript.

Reviewer #1— Minor comment#2:

They should start a new paragraph starting from sentence starting on line 213.

Response:

We have made this change in our manuscript.

Reviewer #1— Minor comment#3:

Figure 3C: I propose that the panels for each species be aligned vertically to help the reader

visualize inter-specific variation in the distribution of I-III type transitions. This would also allow one to map these panels onto the phylogeny (which the authors already have by sorting them horizontally according to phylogenetic relatedness). It would also be neat to compare a dendrogram based on sequence transitions (e.g., cluster analysis) vs. that based on genetic data (i.e., phylogeny).

Response:

We have revised Figure 3C based on the reviewer's suggestion.

Reviewer #1— Minor comment#4:

Figure 4A,B: Why aren't Type III transitions plotted in the figure describing developmental changes? ZF and BFs are quite distinct in the frequencies of these transitions as well (Figure 3).

Response:

Based on this suggestion, we added information related to Type III transitions in our revised Figures 4A and B.

Reviewer #1— Minor comment#5:

Figure 4A,B: spectrograms are too small. I suggest breaking this figure up into two figures, one for each species, and enlarging the top panels (e.g., A+C and B+D). (What is currently Figure 4E can be a stand alone figure)

Response:

We have increased the size of Figure 4 and renumbered Figure 4E as Figure 5 in our revised manuscript.

Reviewer #1— Minor comment#6:

Figure 4C, 4D: I question the utility of the empty circles depicting "BF adults" and "ZF adults". Only the tutor is important here, if I understand things correctly. In addition, while I can see why the authors used the same axes for ZF and BF song analysis, it is difficulty to see each individual bird's trajectory. For example, because the y-axis ranges from 0-15% but the values for ZF birds only range between 0-5%, you don't see much of the variation in that axis. That might be the point, but overall those panels are small and difficult to interpret. Another option is to just expand the entire figure.

Response:

The purpose of Figures 4C and D was to represent each individual bird's trajectory of song development against tutored songs and to show the development of species-specificity of the song patterns. Therefore, we showed empty circles depicting "BF and ZF adults" to indicate the typical ranges of species-specificity of ZF and BF along the same axes indicating the transition types I

and II. To address the difficulties in properly visualization each individual bird's trajectory, we expanded the entire figures, as suggested by the reviewer.

Reviewer #2— Summary:

The manuscript presents a novel method for quantitative characterization of birdsong syntax in a manner that allows comparative analysis of different species, as well as assessment of inter-individual variability within a species and of vocal development. I think the method is likely to prove very useful, and complements nicely existing methods. It is therefore definitely suitable for publication in the Journal of Nueroscience Methods. What I particularly like about the proposed method is that it does not focus on the specific syllable sequences performed (e.g. ABC vs ACB etc) but on more general features of the bird's syntax, namely its tendency to alternate between different syllable types or to repeat the same type. Such "bird's eye view" of song syntax highlights species-specific syntax traits that may be missed with more specific analyses. The testing of method on data from 6 different species is also a big plus.

I have one comment regarding the method itself, and some minor points regarding presentation and comparisons with existing methods, all of which can be easily addressed.

Response:

We deeply appreciated the comments from the reviewer. Based on the reviewer's suggestions, we have revised the manuscript and added new figures as described below.

Reviewer #2— Comment #1:

A meaningful description of syntax depends on the existence of distinct types of behavioral elements, in the case of birdsong, on the performance of distinct and stereotyped syllable types. When syllable performance is very variable across individual renditions (for example in juvenile vocalizations), any division into types (A, B etc) becomes arbitrary, and therefore meaningless. This is indeed a serious problem with manual categorization of syllables, which the authors rightly strive to avoid. However, the SSM method does not absolutely solve this problem. It still lurks in their binarization of similarity scores of 2x2 cells. The authors set (arbitrary?) thresholds of 0.86 and 0.33 % similarity as indicating two similar or two different syllables, respectively, and then choose the closest transition type for a given 2x2 cell (supplementary material). This works well for adult individuals belonging to a species with stereotyped performance of syllable types, because in such cases, any cell will be very close to one of the 12 possible types. but it is problematic for the subsong of juvenile birds, and maybe also for adults in a species with variable syllable performance (for example Star finches?). For that reason, I think that parts of the analysis of developmental trajectories are not very meaningful, since it is performed on data that

clearly lacks any crystallized syllable types. For example, the decrease in the performance of type III transitions (fig 4E) is clearly due to developmental crystallization of syllable types, and therefore not really a change in syntax, but in syllable structure. Similarly, the high frequency of type III transitions in star finches might not really be due to longer motifs, but to more variable syllable renditions. I suggest that the authors set some similarity threshold for cell binarization, and restrict their analysis to data where the similarity scores are either high or low (close to black or white), or at least add a discussion of the binarization issue to the manuscript.

Response:

Similarity thresholds used for cell binarization could indeed be arbitrarily chosen by users, depending on their research aims. However, once the threshold is set, it should be kept the same throughout all analyses. This consistency in the threshold value used importantly differs from variability present when human visual inspection is used to categorize syllable types based spectrograms shape.

The initial goal of this study was to detect both intra-species and inter-specific variations in songs as well as developmental changes. Therefore, we first set the similarity threshold strictly to detect different syllables. Under this threshold condition, in the ZF and BF juvenile vocalizations shown in Figure 4A and B, a majority of syllables at around 40–50 phd did not have clear transition rules and were identified as Type III transitions. In contrast, some other syllables in the same song renditions were identified as Type I and II transitions, during the development of species-specific song. We believe that detection of the onset of the species-specificity of vocal patterns and continuous monitoring of the developmental changes would be crucial points in this study. Therefore, as a new quantitative method for detecting species-specific vocal sequence patterns and monitoring developmental dynamics, the SSM method can complement existing methods which are already adapted for quantification of specific syllable transitions. Based on these points, we revised the Discussion section by expanding on points explaining the binarization issue (pages 17 and 18).

To show the effects of different threshold setting for cell binarization, as identified in Reviewer #1's—major concern #2, we have added in our revised manuscript information on the comparison at different threshold settings. Two different threshold settings resulted in similar occurrence rates of the syllable transition types I, II, and III in six of the species, as shown in **Supplementary material 4**.

Reviewer #2— Comment #2:

The authors mention in the introduction previous studies using manual inspection of syllables as a categorization method, but neglect to mention studies using syllable categorization that is not based on manual inspection, but on semi-automatic clustering of syllables based on their mean

acoustic features (Derenocourt et al 2004, Ravbar et al, 2012, Lipkind et al, 2013). This omission should be corrected, and the advantages and disadvantages of the SSM categorization method with respect to other non-manual methods discussed.

Response:

We thank the reviewer for the suggestion. We have cited previous studies using semi-automatic clustering of syllables in the Introduction and Discussion sections (pages 4 and 15). In addition, we revised the Discussion section to explain the uniqueness of the SSM categorization method with respect to existing methods (page 16).

Reviewer #2— Comment #3:

The distinction between quantification of specific transitions and of transition types (motif like, repetitive etc) should be made more clearly, and at an earlier point in the text. This is an important advantage and innovation of the proposed method, but it took me some time, and being well into the methods section, to understand that the authors do not propose a method for quantification of specific syllable transitions, which was previously done.

Response:

We appreciate the reviewer's supportive suggestions on our proposed method. We have added new text to explain the distinction between quantification of specific transitions and transition types in our revised Introduction and Discussion sections (pages 4 and 16).