This is the pre-peer reviewed version of the following article: James et al., Dev. Neurobiol. 2020, which has been published in final form at https://onlinelibrary.wiley.com/doi/epdf/10.1002/dneu.22754.

Manipulations of sensory experiences during development reveal mechanisms underlying vocal learning biases in zebra finches

Logan S. James[1,2,3], Ronald Davies Jr. [1,4], Chihiro Mori[5,6], Kazuhiro Wada[5,7], and Jon T. Sakata[1,2,8*]

1 Department of Biology, McGill University

2 Centre for Research in Brain, Language and Music, McGill University

3 Current address: Department of Integrative Biology, University of Texas at Austin

4 Current address: Department of Computer Science and Engineering, The Ohio State University

5 Graduate School of Life Science, Hokkaido University

6 Current address: Graduate School of Arts and Sciences, University of Tokyo

7 Faculty of Science, Hokkaido University

8 Center for Studies of Behavioral Neurobiology, Concordia University

Running title: Mechanisms of learning biases

*corresponding author

Jon T. Sakata, PhD

Jon.sakata@mcgill.ca

Acknowledgements

We thank M. Kao and S. Kojima for recordings of isolate zebra finches from the University of California, San Francisco, and S.C. Woolley and R. Krahe for input on data analysis and manuscript preparation. This work was supported by funding from the National Science and Engineering Research Council (#05016 to J.T.S.); Fonds de Recherche Nature et Technologies (2018-PR-206494), MEXT/JSPS KAKENHI (JP17H06380 and JP19H04888 to K.W.), the Centre for Research on Brain, Language and Music (L.S.J.), and a Heller award (L.S.J.). The authors declare no conflict of interest.

Abstract

Biological predispositions in learning can bias and constrain the cultural evolution of social and communicative behaviors (e.g., speech and birdsong), and lead to the emergence of behavioral and cultural "universals". For example, surveys of laboratory and wild populations of zebra finches (*Taeniopygia guttata*) document consistent patterning of vocal elements ("syllables") with respect to their acoustic properties (e.g., duration, mean frequency). Furthermore, such universal patterns are also produced by birds that are experimentally tutored with songs containing randomly sequenced syllables ("tutored birds"). Despite extensive demonstrations of learning biases, much remains to be uncovered about the nature of biological predispositions that bias song learning and production in songbirds. Here we examined the degree to which "innate" auditory templates and/or biases in vocal motor production contribute to vocal learning biases and production in zebra finches. Such contributions can be revealed by examining acoustic patterns in the songs of birds raised without sensory exposure to song ("untutored birds") or of birds that are unable to hear from early in development ("early-deafened birds"). We observed that untutored zebra finches and early-deafened zebra finches produce songs with positional variation in some acoustic features (e.g., mean frequency) that resemble universal patterns observed in tutored birds. Similar to tutored birds, early-deafened birds also produced song motifs with alternation in acoustic features across adjacent syllables. That universal acoustic patterns are observed in the songs of both untutored and early-deafened birds highlights the contribution motor production biases to the emergence of universals in culturally transmitted behaviors.

Keywords: birdsong, songbird, biological predispositions, cultural evolution, behavioral universals

Introduction

Biological predispositions can sculpt the acquisition of many important behaviors. For example, the process of filial imprinting is critical for young animals to learn the identity of their caregivers, and young chicks are predisposed to imprint on visual stimuli that resemble adult conspecifics (Sluckin and Salzen, 1961; Horn and McCabe, 1984; Bolhuis and Trooster, 1988; Di Giorgio et al., 2017). Relatedly, songbirds learn their songs during development, and many species of songbirds demonstrate significant biases in learning conspecific song over heterospecific song (reviewed in Marler, 1970a, 1997; Doupe and Kuhl, 1999; Catchpole and Slater, 2008; Wheatcroft and Qvarnström, 2015). Internal biases have also been proposed to guide the acquisition of human music, speech, and language and to contribute to "universal" patterns in these behaviors across cultures (Patel, 2008; Yang et al., 2017; Mehr et al., 2018). Indeed, controlled experiments provide support for a causal role of learning biases (separate from cultural propagation) in the acquisition and transmission of universal behavioral patterns in humans (Ravignani et al., 2016; Kirby, 2017; Culbertson and Schuler, 2019; Ferdinand et al., 2019) and songbirds (Gardner et al., 2005; Fehér et al., 2009; Plamondon et al., 2010; James and Sakata, 2017). However, despite numerous examples of learning biases, developmental processes and mechanisms (e.g., motor and/or sensory contributions) that could underlie biases in learning and transmission are not fully understood.

Songbirds provide a powerful model system to not only reveal predispositions in learning but also to discover mechanisms that contribute to learning biases (Marler and Peters, 1988; Gardner et al., 2005; Fehér et al., 2009; Plamondon et al., 2010). For example, when juvenile zebra finches (*Taeniopygia guttata*) are individually tutored with randomized and unbiased sequences of species-typical vocal elements ("syllables"), they produce convergent songs with common acoustic patterns (James and Sakata, 2017). In particular, zebra finches tutored in this manner ("tutored birds") produce songs in which the syllable positioned at the end of the "motif" (i.e., the primary sequence that is repeated throughout the song) tends to be the longest in duration and in which syllables positioned in the middle of the motif tend to be higher in mean frequency. In addition, tutored birds sequence their syllables such that mean frequency alternates across adjacent syllables within the motif. Interestingly, both the positional variation and alternation in acoustic structure in the songs of these experimentally-tutored birds mirror patterns observed across populations of normally-tutored zebra finches (Zann, 1996; Lachlan et al., 2016; James and Sakata, 2017). Moreover, these acoustic patterns in zebra finches (as well as various acoustic patterns in other songbird species) resemble universal patterns observed in human speech and music, including utterance-final elongation, harmonic arches in phrase structure, and frequency alternation (Tierney et al., 2011; Mol et al., 2017).

Birdsong is learned during a critical period in development, and song learning consists of both sensory and sensorimotor learning (reviewed in Brainard and Doupe, 2002; Tschida and Mooney, 2012; Ikeda et al., 2020; Sakata and Yazaki-Sugiyama, 2020). Briefly, song learning involves memorizing the structure and sequencing of sounds that juvenile songbirds are exposed to during the critical period for song learning, followed by a protracted period of sensorimotor learning in which young birds use auditory feedback and reinforcement learning to shape their vocalizations into songs that match the memorized songs (reviewed in Konishi, 2004; Fee and Goldberg, 2011; Brainard and Doupe, 2013; Murphy et al., 2017, 2020). Song acquisition could be influenced by intrinsic sensory mechanisms that bias, for example, which sounds are preferentially attended to and memorized and that lead to the emergence of song universals. Additionally, central and peripheral aspects of the vocal motor system that bias the production of particular types of sound patterns could similarly sculpt developmental learning and lead to song universals (Riede and Goller, 2014; Ikeda et al., 2020; reviewed in Sakata and Yazaki-Sugiyama, 2020).

A powerful approach to reveal intrinsic sensory or motor biases that shape vocal production and learning is to analyze and compare the songs of birds that have not been exposed to song during the critical period for song learning ("untutored birds") or of birds that were deafened early in development ("early-deafened birds"; e.g., Thorpe, 1954, 1958; Konishi, 1965b, 1965a; Immelmann, 1969; Price, 1979; Marler and Sherman, 1983, 1985; Eales, 1985, 1989; Scharff and Nottebohm, 1991; Kagawa et al., 2014). Songbirds are hypothesized to possess an "innate template" for song that can, for example, underlie tuning and preferences for conspecific song over heterospecific song (Dooling and Searcy, 1980; Marler and Peters, 1988; Nelson and Marler, 1993; Podos, 1997, 1996; Whaling et al., 1997; reviewed in Marler, 1997; Braaten and Reynolds, 1999; Podos et al., 2004; Gardner et al., 2005; Lahti et al., 2011; Prather et al., 2012; ter Haar et al., 2014; Wheatcroft and Qvarnström, 2015, 2017; Araki et al., 2016; Wang et al., 2019; Moore and Woolley, 2019; reviewed in Sakata and Yazaki-Sugiyama, 2020), and the adult songs of untutored birds are thought to reflect interactions between this "innate template", auditory feedback processing, and vocal production biases (Marler, 1997; Konishi, 2004; Soha, 2017; Love et al., 2019). In addition, because birds that are deafened before the onset of the critical period for song development ("early-deafened birds") cannot rely on auditory feedback to shape song development, the songs of early-deafened birds are thought to primarily reflect motor biases (central and peripheral) in song production that could influence vocal learning processes (Thorpe, 1958; Konishi, 1964, 1965b; Nottebohm, 1968; Price, 1979; Suthers and Goller, 1997; Marler, 1997; Larsen and Goller, 1999, 2002; Podos et al., 2009; Riede and Goller, 2014; Düring and Elemans, 2016). While the nature of motor biases remains unclear, such motor biases could reflect activity within "innate" central pattern generators for song production (see Konishi, 2004, 2010).

A number of studies have documented similarities in the acoustic and temporal structures of song between normally-raised birds, untutored birds, and early-deafened birds and provided support for sensory and motor contributions to species-typical song features (reviewed in Marler, 1997; Konishi, 2004, 2010). For example, although untutored and early-deafened zebra finches produce songs in which syllable durations are more variable and in which inter-syllable gaps are longer, they produce songs that are broadly patterned in a manner similar to normally-tutored zebra finches (Price, 1979; Eales, 1985; Scharff and Nottebohm, 1991; Kojima and Doupe, 2009; Mori and Wada, 2015); for example, like normally-tutored birds, both untutored and early-deafened zebra finches produce songs that consist of a repetition of a single stereotyped sequence of syllables. Similarly, comparative studies of swamp and song sparrows document that species differences in repertoire size and in gap and phrase durations are also found among untutored and early-deafened sparrows (Marler, 1967, 1997; Marler and Sherman, 1983, 1985), and untutored cardinals produce songs that retain a number of temporal features of normal cardinal song, including syllable repetition (Dittus and Lemon, 1969).

However, the degree to which the songs of untutored or early-deafened zebra finches are patterned in a manner that matches universal acoustic patterns remains unknown. Specifically, it is not known whether the motifs of untutored or early-deafened zebra finches contain the universal characteristics of utterance-final elongation, harmonic arches, or frequency alternation (Tierney et al., 2011; Lachlan et al., 2016; James and Sakata, 2017). To address this, we examined acoustic patterning within the song motifs of untutored and early-deafened birds and speculate on how such patterning reflects the contributions of sensory and motor biases to song development.

Methods

Animals and song recording

We compared acoustic patterning within the songs of zebra finches (*Taeniopygia guttata*) that remained untutored throughout development ("untutored birds"), were deafened early in development ("early-deafened birds"), or were tutored with randomized and unbiased sequences of syllables ("tutored birds"; see below). All animal experiments were approved by the McGill University Animal Care and Use Committee in accordance with the guidelines of the Canadian Council on Animal Care (McGill University, #7149) or the Committee on Animal Experiments (Hokkaido University, #18-0053) in accordance with Japanese regulations for animal welfare (e.g., Law for the Humane Treatment and Management of Animals with partial amendment No.105, 2011).

Tutored birds: Birds for tutoring experiments were raised by both parents in a sound-attenuating chamber up to 5 days of age, at which time their father was removed. Thereafter, the juvenile was raised by only their mother. Because only male zebra finches learn to produce complex songs and because the critical period for song learning starts ~20 days post-hatching (Roper and Zann, 2006; Brainard and Doupe, 2013), this protocol ensured that juveniles were not exposed to song from a live adult during the critical period of song learning and were naïve to song before experimental tutoring. When these juveniles were nutritionally independent (~30-40 days old), they were housed individually in a sound-attenuating chamber for song tutoring until they were 4 months of age. The birds analyzed here are the same as those examined in James and Sakata (2017).

Juveniles were individually and operantly tutored with synthesized and randomized sequences of five canonical zebra finch syllables (see James and Sakata, 2017 for further description of stimuli). Zebra finch song bouts consist of a repetition of a single sequence of syllables called a "motif", and birds were experimentally tutored with motif (sequence) variants that each contained each of the five syllables only once, a feature that is typical of zebra finch song. Song bouts were synthesized such that each bout contained four motifs, with each motif being a distinct variant (i.e., same syllables arranged in different sequences), and each bird was tutored with every motif variant (n=120 motif variants with a different five-syllable sequence; Figure 1A,B). Each of the five syllable types was a single token taken from a different bird. To ensure unpredictability of tutoring, the sequence of motif variants per song bout was randomized across days for each bird and across birds. Moreover, to ensure that each motif variant was heard equally often, stimuli were scheduled for presentation in blocks of 120 motifs so that every motif variant was heard once before any variant was repeated. Because syllable sequencing is randomized and unbiased in this tutoring regime, this process can reveal biological predispositions in vocal sequence learning.

The timing of syllables and motif variants in song stimuli followed a species-typical pattern: syllables within the motif were each separated by a silent gap of 30 ms, whereas the gap between syllables across adjacent motifs (i.e., from offset of last syllable in the motif to onset of first syllable of subsequent motif) was 100 ms.

Tutored birds were operantly tutored using perch hops for >1 month (ten Cate, 1991; Adret, 1993; Mets and Brainard, 2018, 2019). Song playback was operantly triggered by perch hops using custom-built perches connected to a National Instruments PCI-6503 I/O card (National Instruments, TX), with each perch hop triggering the playback of one song bout. Song playbacks were spaced in time such that juveniles could hear only 10 operantly triggered song playbacks (i.e., 40 motif variants) within each of three time periods in the day (morning, noon, and afternoon). Sound Analysis Pro 2011 (SAP; http://soundanalysispro.com) was used for song tutoring, and stimuli were played out of an Avantone Pro Mixcube speaker (Avantone, NY) connected to a Crown XLS 1000 amplifier (Crown Audio, IN). Acoustic patterning within the adult songs of tutored birds (4 months old) that contained readily identifiable "motifs" was analyzed (n=45 birds; see below for definition).

Untutored birds: The adult songs of untutored birds are thought to reflect developmental interactions between an "innate" song template, auditory feedback processing, and vocal production biases. We analyzed the songs of untutored zebra finches that produced readily identifiable motifs. Untutored birds were from multiple lab populations (n=14 in total; 10 from the University of California, San Francisco, 2 from Hokkaido University, and 2 from McGill University; see below for motif identification). The songs of some of these birds were included in previous studies but acoustic patterning in their songs was not analyzed (Kojima and Doupe, 2007; Mori and Wada, 2015; Chen et al., 2016). Untutored birds were raised by both parents in a sound-attenuating chamber up to 9.2 ± 0.5 (mean ± SEM) days of age (range: 5-12 days), at which point the father was removed and juveniles were raised by only their mother; consequently, the father was removed well before the opening of the critical period for song learning. When these juveniles could feed themselves (~30-40 days post-hatch), they were housed individually throughout development until their songs were recorded in adulthood. Most (11 out of 14) of the untutored birds were recorded when they were 4-5 months old. However, two of the three remaining birds were recorded at 11-12 months of age, and the exact age of the third bird was unknown but he was recorded sometime between 5-12 months of age.

Early-deafened birds: Whether they are tutored or not, juveniles engage in vocal practice during development, and the act of hearing one's own song during vocal practice shapes the nervous system and vocal patterning (reviewed in Marler, 1997; Brainard and Doupe, 2000; Konishi, 2004; Tschida and Mooney, 2012). Therefore, to evaluate audition-independent contributions to acoustic patterning, we also analyzed acoustic patterns within the songs of juvenile zebra finches that were deprived of auditory feedback from early in development, before the opening of the critical period for song learning ("early-deafened birds"; Mori and Wada, 2015). These birds were deafened before fledging (17–23 days post-hatch) by bilateral cochlear extirpation as described previously (for details on experimental procedure: Mori and Wada, 2015). After bilateral cochlear removal, birds were returned to their nest until 32–41 days post-hatch and then housed in group cages with other deafened birds. An important requirement of our analysis is that birds produce a stereotyped motif, and while zebra finches that are deafened early in development eventually produce stereotyped motifs, it takes longer for them to exhibit stereotyped vocal patterns (Mori and Wada, 2015). We analyzed acoustic patterning within the motifs of early-deafened birds that were 11.7 ± 1.5 (mean ± SEM) months old (n=8 birds). The songs of early-deafened birds were described in previous publication dealing with the neurogenomic aspects of song development (Mori and Wada, 2015) but no analyses of acoustic patterns within motifs have been conducted.

All birds (n=67 total across all treatment groups) were individually housed in sound-attenuating chambers during song recording (i.e., only undirected songs were analyzed) and were recorded using Sound Analysis Pro (SAP).

Identifying song motifs

We analyzed the adult songs of tutored, untutored, and early-deafened birds. We identified and labeled syllables and motifs following amplitude-based segmentation of audio files using custom-written MATLAB scripts (mean±SEM: 107.4 ± 7.8 motifs per bird; range: 22 – 285). The song of an individual zebra finch generally consists of a single stereotyped sequence of syllables called a "motif". An individual's motif is readily identifiable because it is repeated multiple times within a song bout (Scharff and Nottebohm, 1991; ten Cate, 1991; Zann, 1996; Tchernichovski et al., 2000) and because the duration of gaps between motifs is usually longer and more variable than gap durations within the motif (Vu et al., 1994; Glaze and Troyer, 2006). To identify motifs and birds for analyses, two authors (LSJ & JTS) independently examined multiple renditions of each bird's song and identified motifs based on sequence repetition, sequence stereotypy and gap durations. Birds were included in the analysis only if both researchers identified and agreed upon a stereotyped motif (n=45 tutored, n=14 untutored, and n=8 early-deafened birds; Figure 1C-E).

Analysis of acoustic features

We analyzed variation in the acoustic features of syllables according to their position in the motif (beginning, middle, or end) as well as changes in these acoustic features across adjacent syllables; universals in such acoustic patterning have been observed in wild and laboratory populations of zebra finches as well as in human speech and music (Zann, 1996; Tierney et al., 2011; Lachlan et al., 2016; James and Sakata, 2017). We measured the duration, mean frequency, mean amplitude, spectral entropy, spectrotemporal entropy, and amplitude entropy (see Supplementary Information for feature definitions) of each rendition of each syllable in every bird's motif, and then analyzed positional variation and alternation in the medians of these features (James and Sakata, 2017). Together, these features have been demonstrated to differ between syllable types, to change across development and with experimental manipulations of the nervous system, to be encoded by neurons in the auditory system, and/or to be regulated by neural activity and morphology of the vocal motor system (e.g., Tchernichovski et al., 2001; Hampton et al., 2009; Stepanek and Doupe, 2010; Wohlgemuth et al., 2010; Piristine et al., 2016; Moore and Woolley, 2019; Isola et al., 2020). For the analysis of positional variation, we considered the first syllable of the motif as the "beginning syllable", the last syllable of the motif as the "end syllable", and syllables between the beginning and end syllables as "middle syllables".

Machine learning techniques

We employed a machine learning algorithm to simultaneously analyze all of the above acoustic features and reveal the feature(s) that provided the most predictive information about syllable position in a bird's motif. We used random forests (RFs) to classify each syllable within renditions of a bird's motif as beginning, middle or end syllables based on their acoustic features. We conducted RFs in R with the randomForest package (Liaw and Wiener, 2002), using the default settings with the exception that we set the number of trees to 250. Prior to training on the algorithms, we normalized the data for each acoustic feature by subtracting the mean and dividing by the standard deviation. We performed a 10-fold validation of the RF model to get a reliable assessment of performance. For this, we first randomized the entire dataset for an individual bird (i.e., all motifs for the bird) and then divided each dataset into 10 sets. We then trained the RF on nine of the sets (i.e., 90% of the data) and tested the accuracy of the trained algorithm against the withheld set ("test set"; i.e., 10% of the data). We repeated this process 10 times such that each of the 10 sets was used once as the test set. Performance of the algorithm across test sets was averaged across the 10 iterations and reported in the manuscript.

To determine the contribution of each feature to the categorization of position (for each bird), we extracted the "importance score" of each feature in the RF algorithm. The importance score represents the efficacy of a particular feature at classifying data (the greater the efficacy of the feature, the higher the importance score). This score is measured in arbitrary units; therefore, for our statistical analyses, we normalized the importance scores for each bird by dividing each importance score by the maximum importance score for that bird (i.e., "relative importance score", in which the most important feature for that bird is scored as a "1"). We computed relative importance scores for each feature during each iteration of the 10-fold validation and report the average relative importance across the iterations.

Statistical analyses

We used linear mixed-effects models (LMMs) using the R package lme4 (Bates et al., 2015) to analyze all data dealing with positional variation and relative importance scores. The data for duration and mean amplitude were highly right-skewed, so we log-transformed those data to improve normality. All three measures of entropy were highly left-skewed, and to normalize these data, we used a reflected log, which is defined as:

$$\text{transformedEntropy}_i = -\log(K - \text{Entropy}_i)$$

where "K" is defined as the maximum entropy in the dataset plus 1 and "i" refers to the ith data point (Tabachnick et al., 2007).

For the main analyses of positional variation of acoustic features, we ran separate LMMs for each acoustic feature and each experimental group (tutored, untutored, or early-deafened bird) with Position (beginning, middle or end) as the independent variable and individual bird ID as a random variable (to account for the fact that multiple syllables were analyzed per bird). We report the results from the analysis of each individual acoustic feature because a preliminary full-factorial mixed effects model with

Position and Feature as independent variables (and bird ID as a random variable) indicated significant differences in positional variation across song features ($p<0.0001$ for Position x Feature interaction for each group of birds).

Similar mixed effects models were used to analyze variation in relative importance scores across features. For the main analyses of importance scores within each group of birds, Feature was the sole independent variable and bird ID was the random variable.

For all mixed-effects models, a type II Wald chi-square test using the R package 'car' was used to determine the significance of each term (Fox et al., 2013), and Tukey's HSD tests with the Holm correction were used for post-hoc contrasts with the R package 'multcomp' (Hothorn et al., 2008).

We also conducted Monte Carlo simulations for two sets of analyses: (1) to assess similarities among experimental groups in how song features change across positions and (2) to assess the extent to which acoustic features alternated within each bird's motif (James and Sakata, 2017). To assess group similarities in the magnitude of positional variation (see Results), we first computed the percent change in mean values (for each acoustic feature within each experimental group) across all pairwise contrasts for position (beginning vs. middle, beginning vs. end, middle vs. end; e.g., percent change in mean frequency from beginning to end syllables for tutored birds). We did this independently for each experimental group, and this calculation yielded 18 percent change values per experimental group (three changes for each of the six features). We analyzed the relationship in corresponding percent changes values between two experimental groups (i.e., the percent change of a spectral feature in one group as the predictor and the percent change for the same acoustic feature in another group as the response variable) and then computed the r^2 of the relationship ("observed r^2"). Because percent changes are not necessarily independent of each other, we assessed the significance of each observed r^2 value using Monte Carlo simulations. For this, we iteratively randomized the positional assignment (beginning, middle or end) of each mean feature value for each experimental group and then recomputed the percent changes. We then calculated the r^2 value of the relationship for the iteration, and repeated the randomization 10,000 times to obtain a distribution of r^2 values expected by chance if positional variation was randomized. To obtain a p-value associated with the observed r^2, we calculated the percent of randomizations that produced an r^2 value greater than or equal to the observed r^2.

For the analysis of alternation, we first computed, within each bird, the observed differences in the median value of an acoustic feature across adjacent syllables (James and Sakata, 2017). We plotted the change in an acoustic feature from the previous $\text{syllable}_{(t-1)}$ to the current $\text{syllable}_{(t)}$ against the change from the current $\text{syllable}_{(t)}$ to the next $\text{syllable}_{(t+1)}$ across all birds to measure the extent to which the feature alternated between adjacent syllables in the motif. If there is acoustic alternation from syllable-to-syllable, we would expect a negative relationship between the change from $\text{syllable}_{(t-1)}$ to $\text{syllable}_{(t)}$ and from $\text{syllable}_{(t)}$ to $\text{syllable}_{(t+1)}$. For each acoustic feature, we then computed the r^2 value of the relationship between the two changes ("observed r^2"). Finally, we compared this observed r^2 value to the distribution of r^2 values obtained from Monte Carlo simulations (10,000 iterations) in which the sequence of syllables within each bird's motif was randomized during each iteration. We computed the percent of r^2 values that were equal to or greater than the observed r^2 to obtain a p-value for the observed r^2. These analyses were conducted separately in each group of birds. All simulations were run in MATLAB.

Results

Positional variation in features

A previous study demonstrated that birds tutored with randomized sequences produced motifs in which the beginning syllable (the first syllable of the motif) and end syllable (the last syllable in the motif) are lower in pitch than middle syllables (syllables between the beginning and end syllables; Figure 2A), and in which the end syllable is longer than other syllables in the motif (Figure 2A-B) (James and Sakata, 2017)(see also Zann, 1996; Lachlan et al., 2016). We extended these published analyses by examining variation in measures of entropy (spectral, spectrotemporal, and amplitude entropy; see Methods, Supplementary Information). There was significant variation across positions in amplitude entropy (LMM: χ^2_2=34.2, p<0.0001; Figure 2D), with end syllables being higher in amplitude entropy than either middle (Tukey's HSD: p<0.0001) or beginning (p=0.0001) syllables, and in spectrotemporal entropy (χ^2_2=21.3, p<0.0001; Figure 2E), with middle syllables being lower in spectrotemporal entropy than either end (p<0.0001) or beginning syllables (p=0.0147). There was no significant positional variation in the spectral entropy of syllables for tutored birds (χ^2_2=4.8, p=0.0906; Figure 2F).

Within the songs of untutored birds, the mean frequency of syllables significantly varied across position (χ^2_2=9.0, p=0.0114; Figure 2G), with middle syllables being significantly higher in mean frequency than beginning syllables (p=0.0336) and marginally higher than end syllables (p=0.0598). There was also significant positional variation in amplitude entropy in the songs of untutored birds (χ^2_2=12.1, p=0.0024; Figure 2J), with end syllables being higher in amplitude entropy than either middle (p=0.0025) or beginning (p=0.0115) syllables, as well as significant variation in spectrotemporal entropy (χ^2_2=13.4, p=0.0012; Figure 2K), with end syllables being higher in spectrotemporal entropy than either middle (p=0.0013) or beginning (p=0.0065) syllables. We did not detect significant variation across positions in duration (χ^2_2=4.3, p=0.1186), amplitude (χ^2_2=1.7, p=0.4184), or spectral entropy (χ^2_2=2.2, p=0.3265) in the songs of untutored birds.

The motifs of early-deafened birds also demonstrated some positional variation. The mean frequency of syllables significantly varied across positions (χ^2_2=10.4, p=0.0055; Figure 2M), with middle syllables being significantly higher in mean frequency than beginning syllables (p=0.0045) and with end syllables being marginally higher than beginning syllables (p=0.0662). Amplitude also significantly varied across positions (χ^2_2=7.3, p=0.0254; Figure 2O), with beginning syllables being significantly lower in amplitude than either middle (p=0.0478) or end (p=0.0478) syllables, and syllable duration demonstrated marginally significant variation across positions (χ^2_2=4.8, p=0.0890), with end syllables tending to be longer than middle and beginning syllables. There was no significant positional variation in amplitude entropy (χ^2_2=1.5, p=0.4760) or spectrotemporal entropy (χ^2_2=2.4, p=0.2972) for early-deafened birds.

Many of the patterns observed in the motifs of untutored birds and early-deafened birds qualitatively resembled patterns observed in birds that were tutored with randomized sequences of syllables. To statistically assess the extent to which these patterns differed, we analyzed the data from all three groups of birds simultaneously (Group as an independent variable). We did not detect a significant interaction between Group and Position for any feature, suggesting that acoustic patterning was statistically comparable across groups (but see Supplementary Information for variation among Groups). To further assess similarities in positional variation across experimental groups, we analyzed the relationships in the percent changes of song features across motif positions among tutored, untutored, and early-deafened birds (see Methods). If acoustic patterning was similar across experimental groups, then percent changes in acoustic features across motif positions in one group should significantly and positively correlate with corresponding percent changes in another group. These relationships were significant and positive between tutored and untutored birds (Monte Carlo simulations: p=0.0273) and between untutored and early-deafened birds (p=0.0067). The relationship was positive but not statistically significant between tutored and early-deafened birds (p=0.1446).

Acoustic predictors of syllable position

Given that a number of acoustic features significantly varied across motif positions, we used a random forests (RF) algorithm to simultaneously analyze all acoustic features and reveal the feature(s) that provided the most predictive information about syllable position for that bird's song ("relative importance score"; see Methods, Supplementary Information). For tutored birds, we observed significant variation in relative importance scores across features (χ^2_5=191.6, p<0.0001; Figure 4A). Duration was more important than any other feature (e.g., spectrotemporal entropy) at predicting syllable position, while amplitude entropy was the least important feature. For untutored birds, there was a significant effect of Feature (χ^2_5=36.4, p<0.0001) with duration being the most important feature and spectral entropy being the least important. Similarly, there was a significant effect of Feature in early-deafened birds (χ^2_5=34.8, p<0.0001), with duration being the most important feature and amplitude entropy being the least important.

Visual inspection of the distribution of relative importance scores across groups of birds suggested similarities as well as some differences in the relative importance of particular features in predicting syllable position (Figure 4). When we simultaneously analyzed all groups, we observed a significant interaction between Feature and Group (χ^2_{10}=21.9, p=0.0156; as well as a significant main effect of Feature: χ^2_5=239.0, p<0.0001). This significant interaction seems to be driven by variation in the relative importance of spectrotemporal entropy and amplitude entropy between tutored and untutored birds. Indeed, when we repeated this analysis among pairs of experimental groups, we only found a significant interaction between Feature and Group when tutored and untutored birds were analyzed together (χ^2_5=16.5, p=0.0056).

Pairwise sequencing of syllables based on acoustic features

Speech and musical phrases tend to be organized in a manner such that adjacent syllables or notes alternate in acoustic features (e.g., stress patterns in speech; Tierney et al., 2011). We previously reported that syllables in the motifs of tutored zebra finches similarly alternate in mean frequency (Figure 5A) but not in syllable duration or amplitude (James and Sakata, 2017). We extended these analyses by examining three measures of entropy in tutored birds, and observed significant alternation of amplitude entropy (Monte Carlo simulation; p=0.0020; Figure 5B) and spectrotemporal entropy (p=0.0005; Figure 5C) but not of spectral entropy (data not shown). Whereas we did not observe significant alternation for any feature in the songs of untutored birds (p>0.20 for all; Figure 5D-F), we observed marginally significant alternation of mean frequency (p=0.0515; Figure 5G) and significant alternation of amplitude entropy in the songs of early-deafened birds (p=0.0135; Figure 5I).

Discussion

Biological predispositions in learning sculpt the acquisition of many socially important behaviors (Fehér et al., 2009; Ravignani et al., 2016; James and Sakata, 2017; Kirby, 2017; Smith et al., 2017), but the mechanisms that underlie learning biases remain poorly understood. To reveal the extent to which sensory or motor biases could contribute to biases in the acquisition and production of acoustic patterns in zebra finch song, we analyzed the degree to which zebra finches raised without exposure to song during the critical period of song learning ("untutored birds") and zebra finches deafened early in development ("early-deafened birds") produced acoustic patterns that resembled universal patterns observed in tutored birds (Zann, 1996; Lachlan et al., 2016; James and Sakata, 2017). Broadly speaking, we observed that both untutored and early-deafened zebra finches produced songs with acoustic patterning (e.g., positional variation, frequency alternation) that resembled some universal patterns produced by tutored birds, suggesting a contribution of sensory and/or motor biases to the development of these universals.

Similarities in acoustic patterning between tutored birds and untutored or early-deafened birds can provide distinct insights into the factors that shape the acquisition of species-typical patterns. Because untutored birds primarily rely on "innate templates", auditory feedback (but see Suthers et al., 2002), and motor biases for song development, our finding that the songs of tutored and untutored zebra finches (Figures 2 & 3) demonstrate similar patterns of positional variation in mean frequency and amplitude entropy supports the notion that "innate" auditory or motor biases could contribute to universal song patterns. Analyzing the songs of early-deafened birds can provide more specific insight into mechanisms because the songs of early-deafened birds are thought to primarily reflect motor biases in song learning and production (Konishi, 1964, 1965a; Price, 1979; Marler, 1997; Mori and Wada, 2015). Consequently, similarities in positional variation in mean frequency and in the alternation of acoustic features between tutored birds and early-deafened birds provide compelling support for the contribution of motor biases to universal patterns in song (Tierney et al., 2011; Lachlan et al., 2016; James and Sakata, 2017) and support other experiments that highlight motor contributions to temporal patterns of species-typical song (Marler, 1970b, 1997; Price, 1979; Podos, 1996; Gardner et al., 2005;

Podos et al., 2009; Mori and Wada, 2015; Schmidt and Goller, 2016; Alliende et al., 2017). In addition, syllable duration was the best predictor of syllable position within the motifs of all experimental groups of zebra finches, including early-deafened birds, and because respiratory patterns regulate the duration of syllables, we propose that neural, muscular, and/or morphological features that influence respiratory patterns might be particularly influential in generating motor biases that shape vocal learning in zebra finches (Andalman et al., 2011; Schmidt and Goller, 2016). For example, HVC (used as a proper name), the dorsomedial nucleus of the intercollicular complex, and hindbrain respiratory circuits influence the temporal pattern of song (Schmidt and Goller, 2016; Elie and Theunissen, 2020), and interactions between these areas could be critical for creating motor biases that influence song learning.

Despite the numerous similarities in acoustic patterns among tutored, untutored, and early-deafened birds, there exist some differences across groups that can also yield insight into mechanisms of song development. Differences in acoustic patterning between the songs of untutored birds and early-deafened birds are particularly informative because they can provide insight into “innate” auditory templates (Marler, 1997; Soha, 2017). Both untutored and early-deafened birds possess this innate auditory template but only untutored birds can use it to guide song development. In particular, untutored birds but not early-deafened birds can use auditory feedback to compare their vocalizations to the innate auditory template and change their songs accordingly. In this respect, that untutored birds (and tutored birds) but not early-deafened birds demonstrate positional variation in amplitude entropy and spectrotemporal entropy suggests that interactions between the innate auditory template and auditory feedback could be important for the development of these patterns. It is worth noting, however, that such differences warrant further investigation because sample sizes were smaller for early-deafened birds.

The extent to which similarities in acoustic patterning among tutored, untutored, and early-deafened birds reflect “innate” motor biases remains unknown (and difficult to assess). Sensory experiences *in ovo* and sensory and sensorimotor experiences before the onset of the critical period for song learning could contribute to the emergence of these acoustic patterns. For example, studies have documented effects of *in ovo* auditory experiences on some aspects of vocal development (Colombelli-Négrel et al., 2012; Wheatcroft and Qvarnström, 2017; Katsis et al., 2018). Additionally, juveniles produce a number of unlearned calls, and it is possible that the patterning of such calls shapes neural and vocal development (reviewed in Marler, 2004). While little is known about the temporal patterning of the calls of juvenile and adult zebra finches, it is possible that patterns of call production by juvenile males (e.g., Liu et al., 2009) or patterns of call perception (i.e., hearing calls produced by mothers or female siblings; e.g., Nottebohm, 1968; Price, 1979; Eales, 1987) early in development shaped song development in untutored birds or early-deafened birds. Finally, it is possible that the shared acoustic patterns among tutored, untutored, and early-deafened birds reflect a contribution of somatosensory feedback to vocal development (Suthers et al., 2002), and developmental manipulations of somatosensory feedback are important to reveal this contribution.

While it is possible that early sensory or sensorimotor experiences or non-auditory (e.g., somatosensory) sources of feedback could contribute to the development of song patterns, our data suggest a mechanism that could bias how learned sounds are transformed into species-typical song patterns. Zebra finches that are tutored with sequences that are not species-typical or with stimuli that do not provide reliable information about syllable sequencing converge on universal patterns observed in normally-tutored birds (James and Sakata, 2017). The current results, in particular those from early-deafened birds, suggest that when birds are tutored with songs that lack reliable information about sequencing, the vocal motor neurons in forebrain and hindbrain circuitry will preferentially place sounds with particular acoustic features into positions in the motif that conform with motor biases. Such a model is consistent with the observation that “novel syllables” (i.e., syllables that birds were not tutored with) are sequenced with similar positional variation as imitated syllables in tutored birds (James and Sakata, 2017). More broadly speaking, our findings suggest that motor biases influence when and how specific acoustic patterns emerge over development in normally-reared birds.

Marler (1984, 1997) also speculated about mechanisms underlying the emergence of species-typical universals in birdsong, focusing primarily on sensory biases and templates that could lead to species-typical features of sparrow song (Marler and Sherman, 1983; Marler, 1984, 1997). He proposed that universal acoustic features are manifestations of two types of “innate” templates: “latent templates”, which are activated upon exposure to song, and “preactive templates”, which do not require exposure to song. Under normal circumstances, both types of templates are thought to work in concert, but developmental manipulations can reveal the contribution of each type of template (Soha, 2017). According to this framework, song universals that are present in the adult songs of untutored birds reflect acoustic features encoded by the preactive template; specifically, the adult song of untutored birds are thought to reflect the interaction between the preactive template and auditory feedback. In contrast, song universals that are present in tutored birds but not in the songs of untutored birds are assumed to be encoded by the latent template. Applying this framework to our findings, because untutored zebra finches, like normally-tutored zebra finches (Figures 2 & 3; James and Sakata, 2017), demonstrate harmonic arches and positional variation in amplitude and spectrotemporal entropy, these features are manifestations of the preactive template in zebra finches. However, unlike normally-tutored birds (Figures 2 & 3; James and Sakata, 2017), untutored zebra finches do not demonstrate significant positional variation in syllable duration or alternation in syllable features, suggesting that these song universals could be a reflection of the latent template.

Taken together, these results support the notion that sensory and motor biases contribute to the emergence of universals in culturally-transmitted behaviors. Biological biases shape the cultural transmission of behaviors, including music, language, and birdsong, and this process involves the regularization of sensory inputs by learners (Kirby et al., 2008; Fehér et al., 2009; Tierney et al., 2011; Ravignani et al., 2016; Kirby, 2017; Ferdinand et al., 2019). Our results suggest that the motor system is involved in the regularization processes and that sensory inputs are transformed in a non-random manner to fit biases and constraints of the motor system. This notion is consistent with the "gestural origins theory of speech" (Liberman et al., 1967) and related theories (e.g., frame-content theory) that propose that the building blocks and combinatorial structure of speech are derived from innate articulatory gestures and pre-existing rhythmic behaviors of the jaw (MacNeilage, 1998). Further research in other vocal learning species will be important for understanding how broadly these theories apply to the formation of universal vocal patterns across species (Murphy et al., 2017).

Data availability statement:

Data will be made available on a public repository upon publication of the article.

References

Adret P. 1993. Operant conditioning, song learning and imprinting to taped song in the zebra finch. Anim Behav 46:149–159.

Alliende J, Giret N, Pidoux L, Negro C Del, Leblois A. 2017. Seasonal plasticity of song behavior relies on motor and syntactic variability induced by a basal ganglia–forebrain circuit. Neuroscience 359:49–68.

Andalman AS, Foerster JN, Fee MS. 2011. Control of vocal and respiratory patterns in birdsong: Dissection of forebrain and brainstem mechanisms using temperature. PLoS One 6:.

Araki M, Bandi MM, Yazaki-Sugiyama Y. 2016. Mind the gap: Neural coding of species identity in birdsong prosody. Science 354:1282–1287.

Bates D, Maechler M, Bolker B, Walker S. 2015. Package “lme4.” Convergence 12:.

Bolhuis JJ, Trooster WJ. 1988. Reversibility revisited: Stimulus-dependent stability of filial preference in the chick. Anim Behav 36:668–674.

Braaten RF, Reynolds K. 1999. Auditory preference for conspecific song in isolation-reared zebra finches. Anim Behav 58:105–111.

Brainard MS, Doupe AJ. 2000. Auditory feedback in learning and maintenance of vocal behaviour. Nat Rev Neurosci 1:31–40.

Brainard MS, Doupe AJ. 2002. What songbirds teach us about learning. Nature 417:351–358.

Brainard MS, Doupe AJ. 2013. Translating birdsong: songbirds as a model for basic and applied medical research. Annu Rev Neurosci 36:489–517.

Catchpole CK, Slater PJB. 2008. Bird Song: Biological Themes and Variations. Cambridge, UK: Cambridge University Press.

Cate C ten. 1991. Behaviour-contingent exposure to taped song and zebra finch song learning. Anim Behav 42:857–859.

Chen Y, Matheson LE, Sakata JT. 2016. Mechanisms underlying the social enhancement of vocal learning in songbirds. Proc Natl Acad Sci 113:6641–6646.

Colombelli-Négrel D, Hauber ME, Robertson J, Sulloway FJ, Hoi H, Griggio M, Kleindorfer S. 2012. Embryonic learning of vocal passwords in superb fairy-wrens reveals intruder cuckoo nestlings. Curr Biol 22:2155–2160.

Culbertson J, Schuler K. 2019. Artificial language learning in children. Annu Rev Linguist 5:353–373.

Dittus WPJ, Lemon RE. 1969. Effects of song tutoring and acoustic isolation on the song repertoires of cardinals. Anim Behav 17:523–533.

Dooling R, Searcy M. 1980. Early perceptual selectivity in the swamp sparrow. Dev Psychobiol 13:499–506.

Doupe AJ, Kuhl PK. 1999. Birdsong and humman speech: Common themes and mechanisms. Annu Rev Neurosci 22:567–631.

Düring DN, Elemans CPH. 2016. Embodied motor control of avian vocal production. Vertebrate Sound Production and Acoustic Communication, Springer, Cham, p 119–157.

Eales LA. 1985. Song learning in zebra finches: some effects of song model availability on what is learnt and when. Anim Behav 33:1293–1300.

Eales LA. 1987. Song learning in female-raised zebra finches: another look at the sensitive phase. Anim Behav 35:1356–1365.

Eales LA. 1989. The influences of visual and vocal interaction on song learning in Zebra finches. Anim Behav 37:507–508.

Elie JE, Theunissen F. 2020. The neuroethology of vocal communication in songbirds: production and perception of a call repertoire. The Neuroethology of Birdsong, Springer, p 175–209.

Fee MS, Goldberg JH. 2011. A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. Neuroscience 198:152–170.

Fehér O, Wang H, Saar S, Mitra PP, Tchernichovski O. 2009. De novo establishment of wild-type song culture in the zebra finch. Nature 459:564–568.

Ferdinand V, Kirby S, Smith K. 2019. The cognitive roots of regularization in language. Cognition 184:53–68.

Fox J, Friendly M, Weisberg S. 2013. Hypothesis tests for multivariate linear models using the car Package. R J 5:39.

Gardner TJ, Naef F, Nottebohm F. 2005. Freedom and rules: the acquisition and reprogramming of a bird’s learned song. Science 308:1046–1049.

Giorgio E Di, Loveland JL, Mayer U, Rosa-Salva O, Versace E, Vallortigara G. 2017. Filial responses as predisposed and learned preferences: Early attachment in chicks and babies. Behav Brain Res 325:90–104.

Glaze CM, Troyer TW. 2006. Temporal structure in zebra finch song: Implications for motor coding. J Neurosci 26:991–1005.

Haar SM ter, Kaemper W, Stam K, Levelt CC, Cate C ten. 2014. The interplay of within-species perceptual predispositions and experience during song ontogeny in zebra finches (*Taeniopygia guttata*). Proc R Soc B Biol Sci 281:20141860.

Hampton CM, Sakata JT, Brainard MS. 2009. An avian basal ganglia-forebrain circuit contributes differentially to syllable versus sequence variability of adult Bengalese finch song. J Neurophysiol 101:3235–3245.

Horn G, McCabe BJ. 1984. Predispositions and preferences. Effects on imprinting of lesions to the chick brain. Anim Behav 32:288–292.

Hothorn T, Bretz F, Westfall P. 2008. Simultaneous inference in general parametric models. Biometrical J 50:346–363.

Ikeda MZ, Trusel M, Roberts TF. 2020. Memory circuits for vocal imitation. Curr Opin Neurobiol 60:37–46.

Immelmann K. 1969. Song development in the zebra finch and other estrildid finches. Bird vocalizations 61–74.

Isola G, Vochin A, Sakata JT. 2020. Manipulations of inhibition in cortical circuitry differentially affect spectral and temporal features of Bengalese finch song. J Neurophysiol.

James LS, Sakata JT. 2017. Learning biases underlie “universals” in avian vocal sequencing. Curr Biol 27:3676-3682.e4.

Kagawa H, Suzuki K, Takahasi M, Okanoya K. 2014. Domestication changes innate constraints for birdsong learning. Behav Processes 106:91–97.

Katsis AC, Davies MH, Buchanan KL, Kleindorfer S, Hauber ME, Mariette MM. 2018. Prenatal exposure to incubation calls affects song learning in the zebra finch. Sci Rep 8:15232.

Kirby S. 2017. Culture and biology in the origins of linguistic structure. Psychon Bull Rev 24:118–137.

Kirby S, Cornish H, Smith K. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. Proc Natl Acad Sci 105:10681–10686.

Kojima S, Doupe AJ. 2007. Song selectivity in the pallial-basal ganglia song circuit of zebra finches raised without tutor song exposure. J Neurophysiol 98:2099–109.

Kojima S, Doupe AJ. 2009. Activity propagation in an avian basal ganglia-thalamocortical circuit essential for vocal learning. J Neurosci 29:4782–4793.

Konishi M. 1964. Effects of deafening on song development in two species of juncos. Condor 66:85–102.

Konishi M. 1965a. The role of auditory feedback in the control of vocalization in the white-crowned sparrow. Z Tierpsychol 22:770–783.

Konishi M. 1965b. Effects of deafening on song development in American robins and black-headed grosbeaks. Z Tierpsychol 22:584–599.

Konishi M. 2004. The role of auditory feedback in birdsong. Ann N Y Acad Sci 1016:463–475.

Konishi M. 2010. From central pattern generator to sensory template in the evolution of birdsong. Brain Lang 115:18–20.

Lachlan RF, Heijningen CAA van, Haar SM ter, Cate C ten. 2016. Zebra finch song phonology and syntactical structure across populations and continents—a computational comparison. Front Psychol 7:980.

Lahti DC, Moseley DL, Podos J. 2011. A tradeoff between performance and accuracy in bird song learning. Ethology 117:802–811.

Larsen ON, Goller F. 1999. Role of syringeal vibrations in bird vocalizations. Proc R Soc London Ser B Biol Sci 266:1609–1615.

Larsen ON, Goller F. 2002. Direct observation of syringeal muscle function in songbirds and a parrot. J Exp Biol 205:25–35.

Liaw A, Wiener M. 2002. Classification and regression by randomForest. R news 2:18–22.

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. 1967. Perception of the speech code. Psychol Rev 74:431–461.

Liu WC, Wada K, Nottebohm F. 2009. Variable food begging calls are harbingers of vocal learning. PLoS One 4:e5929.

Love J, Hoepfner A, Goller F. 2019. Song feature specific analysis of isolate song reveals interspecific variation in learned components. Dev Neurobiol 79:350–369.

MacNeilage PF. 1998. The frame/content theory of evolution of speech production. Behav Brain Sci 21:499–511.

Marler P. 1967. Animal communication signals. Science 157:769–774.

Marler P. 1970a. Birdsong and human speech: could there be parallels? Am Sci 58:669–674.

Marler P. 1970b. A comparative approach to vocal learning: Song development in white-crowned sparrows. J Comp Physiol Psychol 71:1–25.

Marler P. 1984. Song learning: innate species differences in the learning process. The Biology of Learning, Springer Berlin Heidelberg, p 289–309.

Marler P. 1997. Three models of song learning: evidence from behavior. J Neurobiol 33:501–516.

Marler P. 2004. Bird calls: their potential for behavioral neurobiology. Ann N Y Acad Sci 1016:31–44.

Marler P, Peters S. 1988. The role of song phonology and syntax in vocal learning preferences in the song sparrow, *Melospiza melodia*. Ethology 77:125–149.

Marler P, Sherman V. 1983. Song structure without auditory feedback: emendations of the auditory template hypothesis. J Neurosci 3:517–531.

Marler P, Sherman V. 1985. Innate differences in singing behaviour of sparrows reared in isolation from adult conspecific song. Anim Behav 33:57–71.

Mehr SA, Singh M, York H, Glowacki L, Krasnow MM. 2018. Form and function in human song. Curr Biol 28:356-368.e5.

Mets DG, Brainard MS. 2018. Genetic variation interacts with experience to determine interindividual differences in learned song. Proc Natl Acad Sci 115:421–426.

Mets DG, Brainard MS. 2019. Learning is enhanced by tailoring instruction to individual genetic differences. Elife 8:1–15.

Mol C, Chen A, Kager RWJ, Haar SM ter. 2017. Prosody in birdsong: A review and perspective. Neurosci Biobehav Rev 81:167–180.

Moore JM, Woolley SMN. 2019. Emergent tuning for learned vocalizations in auditory cortex. Nat Neurosci 22:1469–1476.

Mori C, Wada K. 2015. Audition-independent vocal crystallization associated with intrinsic developmental gene expression dynamics. J Neurosci 35:878–889.

Murphy K, James LS, Sakata JT, Prather JF. 2017. Advantages of comparative studies in songbirds to understand the neural basis of sensorimotor integration. J Neurophysiol 118:800–816.

Murphy K, Lawley KS, Smith P, Prather JF. 2020. New insights into the avian song system and neuronal control of learned vocalizations. In: Sakata JT, Woolley SC, Fay RR, Popper AN, editors. The Neuroethology of Birdsong, New York, NY: Springer, p 65–92.

Nelson DA, Marler P. 1993. Innate recognition of song in white-crowned sparrows: a role in selective vocal learning? Anim Behav 46:806–808.

Nottebohm F. 1968. Auditory experience and song development in the chaffinch *Fringilla coelebs*. Ibis (Lond 1859) 110:549–568.

Patel A. 2008. Music, Language, and the Brain. Oxford, UK: University Press.

Piristine HC, Choetso T, Gobes SMH. 2016. A sensorimotor area in the songbird brain is required for production of vocalizations in the song learning period of development. Dev Neurobiol 76:1213–1225.

Plamondon SL, Rose GJ, Goller F. 2010. Roles of syntax information in directing song development in white-crowned sparrows (*Zonotrichia leucophrys*). J Comp Psychol 124:117–32.

Podos J. 1996. Motor constraints on vocal development in a songbird. Anim Behav 51:1061–1070.

Podos J. 1997. A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). Evolution (N Y) 51:537.

Podos J, Lahti DC, Moseley DL. 2009. Vocal performance and sensorimotor learning in songbirds.

Advances in the Study of Behavior, p 159–195.

Podos J, Peters S, Nowicki S. 2004. Calibration of song learning targets during vocal ontogeny in swamp sparrows, Melospiza georgiana. Anim Behav 68:929–940.

Prather JF, Peters S, Mooney R, Nowicki S. 2012. Sensory constraints on birdsong syntax: neural responses to swamp sparrow songs with accelerated trill rates. Anim Behav 83:1411–1420.

Price PH. 1979. Developmental determinants of structure in zebra finch song. J Comp Physiol Psychol 93:260–277.

Ravignani A, Delgado T, Kirby S. 2016. Musical evolution in the lab exhibits rhythmic universals. Nat Hum Behav 1:0007.

Riede T, Goller F. 2014. Morphological basis for the evolution of acoustic diversity in oscine songbirds. Proc R Soc B Biol Sci 281:20132306–20132306.

Roper A, Zann R. 2006. The onset of song learning and song tutor selection in fledgling zebra finches. Ethology 112:458–470.

Sakata JT, Yazaki-Sugiyama Y. 2020. Neural circuits underlying vocal learning in songbirds. In: Sakata JT, Woolley SC, Fay RR, Popper AN, editors. The Neuroethology of Birdsong, New York, NY: Springer, p 29–63.

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

Schmidt MF, Goller F. 2016. Breathtaking songs: coordinating the neural circuits for breathing and singing. Physiology 31:442–451.

Sluckin W, Salzen EA. 1961. Imprinting and perceptual learning. Q J Exp Psychol 13:65–77.

Smith K, Perfors A, Fehér O, Samara A, Swoboda K, Wonnacott E. 2017. Language learning, language use and the evolution of linguistic variation. Philos Trans R Soc B Biol Sci 372:20160051.

Soha J. 2017. The auditory template hypothesis: a review and comparative perspective. Anim Behav 124:247–254.

Stepanek L, Doupe AJ. 2010. Activity in a cortical-basal ganglia circuit for song is required for social context-dependent vocal variability. J Neurophysiol 104:2474–2486.

Suthers RA, Goller F. 1997. Motor correlates of vocal diversity in songbirds. Current Ornithology, Boston, MA: Springer US, p 235–288.

Suthers RA, Goller F, Wild JM. 2002. Somatosensory feedback modulates the respiratory motor program of crystallized birdsong. Proc Natl Acad Sci 99:5680–5685.

Tabachnick B, Fidell L, Ullman J. 2007. Using Multivariate Statistics. Boston, MA: Pearson.

Tchernichovski O, Mitra PP, Lints T, Nottebohm F. 2001. Dynamics of the vocal imitation process: How a zebra finch learns its song. Science 291:2564–2569.

Tchernichovski O, Nottebohm F, Ho C, Pesaran B, Mitra P. 2000. A procedure for an automated measurement of song similarity. Anim Behav 59:1167–1176.

Thorpe WH. 1954. The process of song-learning in the chaffinch as studied by means of the sound spectrograph. Nature 173:465–469.

Thorpe WH. 1958. Further studies on the process of song learning in the chaffinch (Fringilla Coelebs Gengleri). Nature 182:554–557.

Tierney AT, Russo FA, Patel AD. 2011. The motor origins of human and avian song structure. Proc Natl Acad Sci 108:15510–15515.

Tschida K, Mooney R. 2012. The role of auditory feedback in vocal learning and maintenance. Curr Opin Neurobiol 22:320–327.

Vu E, Mazurek M, Kuo Y. 1994. Identification of a forebrain motor programming network for the learned song of zebra finches. J Neurosci 14:6924–6934.

Wang H, Sawai A, Toji N, Sugioka R, Shibata Y, Suzuki Y, Ji Y, Hayase S, Akama S, Sese J, Wada K. 2019. Transcriptional regulatory divergence underpinning species-specific learned vocalization in songbirds. PLOS Biol 17:e3000476.

Whaling CS, Solis MM, Doupe a J, Soha J a, Marler P. 1997. Acoustic and neural bases for innate recognition of song. Proc Natl Acad Sci 94:12694–12698.

Wheatcroft D, Qvarnström A. 2015. A blueprint for vocal learning: auditory predispositions from brains to genomes. Biol Lett 11:20150155-.

Wheatcroft D, Qvarnström A. 2017. Genetic divergence of early song discrimination between two young songbird species. Nat Ecol Evol 1:0192.

Wohlgemuth MJ, Sober SJ, Brainard MS. 2010. Linked control of syllable sequence and phonology in birdsong. J Neurosci 30:12936–49.

Yang C, Crain S, Berwick RC, Chomsky N, Bolhuis JJ. 2017. The growth of language: Universal Grammar, experience, and principles of computation. Neurosci Biobehav Rev.

Zann RA. 1996. The Zebra Finch: A Synthesis of Field and Laboratory Studies. Oxford, UK: Oxford University Press.

Figure Legends

Figure 1: Example spectrograms of motifs from the tutoring stimuli and all three groups of birds. Tutoring stimuli (A) consisted of motif variants that contained the same syllables (see colored bars above spectrogram) but differed in sequencing. Two of the 120 possible motif variants are depicted here (see Methods). (B) Circle diagram depicting the overall transition probabilities between syllables in the tutoring stimuli, with lines of the same color depicting transitions from one syllable to the other four and the thickness of the lines corresponding to the transition probability. All lines are of equal width because syllable transitions within the tutoring stimuli are unbiased (James and Sakata, 2017). (C-E) Examples of segments of songs produced by adult zebra finches, with black bars above the spectrogram indicating the motif. This includes examples of (C) a song produced by a bird tutored with random sequences of syllables, (D) a song produced by a bird that remained untutored throughout development and was able to hear (untutored), and (E) a song produced by a bird that was deafened before the critical period for song learning ("early-deafened").

Figure 2. Positional variation of syllable features. Plotted are the features of each syllable produced by tutored birds (A-F), untutored birds (G-L) or early-deafened birds (M-R). Each point depicts the median value for a syllable produced by a bird, and bars indicate the mean ± SEM. Blue bars = tutored birds (n=200 syllables); peach bars = untutored birds (n=61 syllables); green bars = early-deafened birds (n=34 syllables). Amplitude was measured in arbitrary units (see Supplementary Information), and here we depicted the amplitude of syllables normalized to the mean amplitude within each bird's song (i.e., normalized amplitude) to better depict positional variation. "*" indicates contrasts that are significant at $p<0.05$ (Tukey's HSD), "~" indicates contrasts that are marginally significant ($p<0.10$).

Figure 3: Correlations in acoustic changes across motif positions between groups of birds. The mean acoustic change across positions (i.e., from beginning ("beg") to middle ("mid"; triangles), beginning to end (circles), and middle to end (squares)) are plotted for different pairs of groups. Colors indicate each of the six features: mean frequency (light blue), duration (pink), amplitude (dark green), amplitude entropy (dark green), spectrotemporal entropy (brown) and spectral entropy (orange).

Figure 4: A machine learning approach (random forests) reveals the relative importance of acoustic features in predicting syllable position across tutored (A), untutored (B), and early-deafened birds (C). A-C. The relative importance scores for each of the six features within a bird (Dur=duration, STent=spectrotemporal entropy; Freq=mean frequency; Amp=amplitude; Sent=spectral entropy; Aent= amplitude entropy). Each point depicts the relative importance score of that feature for an individual bird. Features that share a letter did not differ significantly in relative importance.

Figure 5: Acoustic changes across adjacent syllables in the motif. The songs of tutored birds exhibited significant alternation of mean frequency (A; James and Sakata, 2017), amplitude entropy (B) and spectrotemporal entropy (C). None of these same features significantly alternated within the songs of untutored birds (D-F) but some features significantly alternated in early-deafened birds (G-I). "*" above panels indicates relationships that are significant at $p<0.05$, "~" indicates relationships that are marginally significant ($p<0.10$).

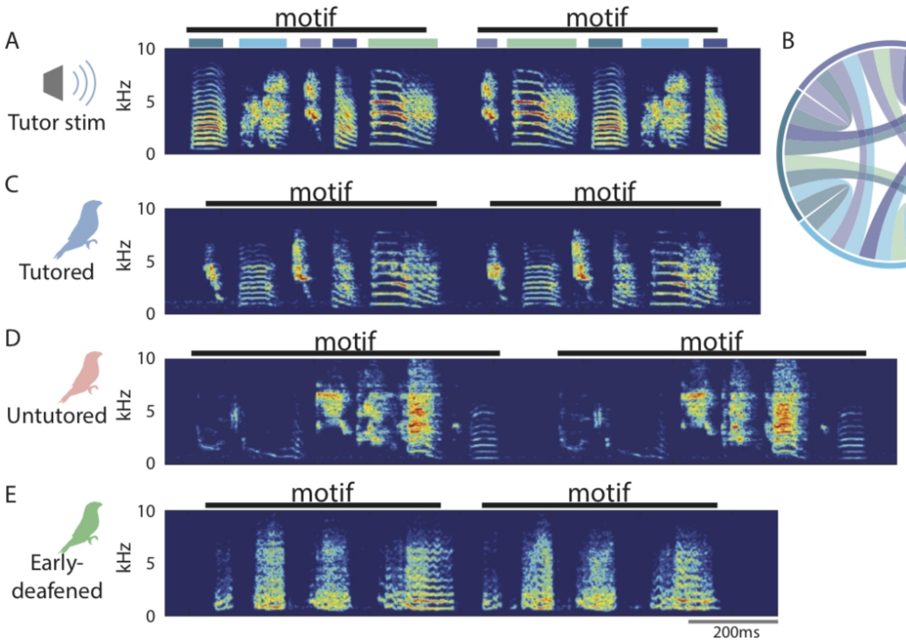

Figure 1

197x126mm (150 x 150 DPI)

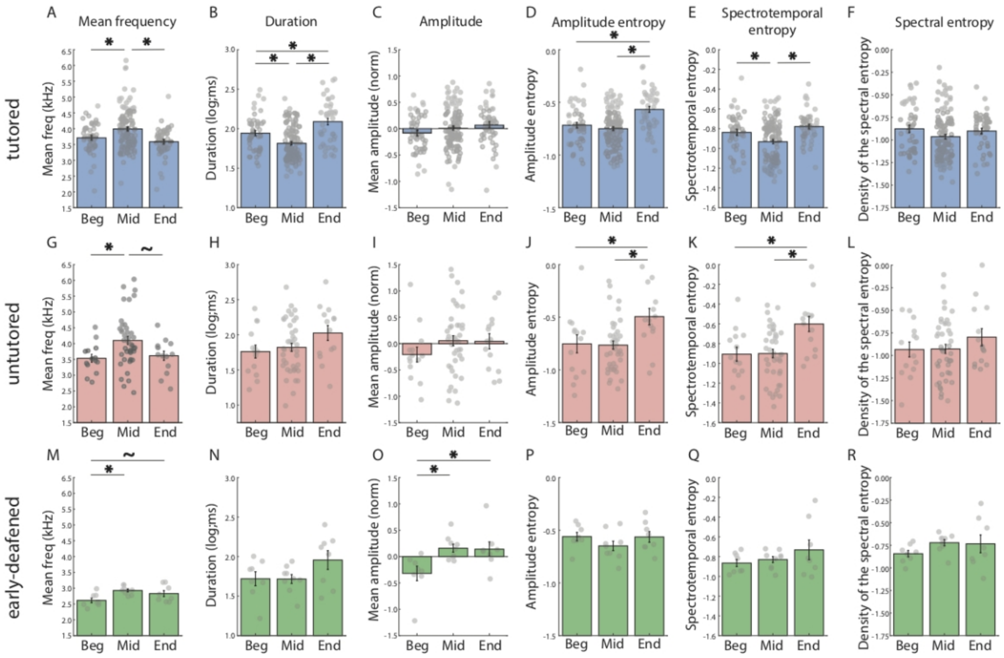

Figure 2

201x131mm (150 x 150 DPI)

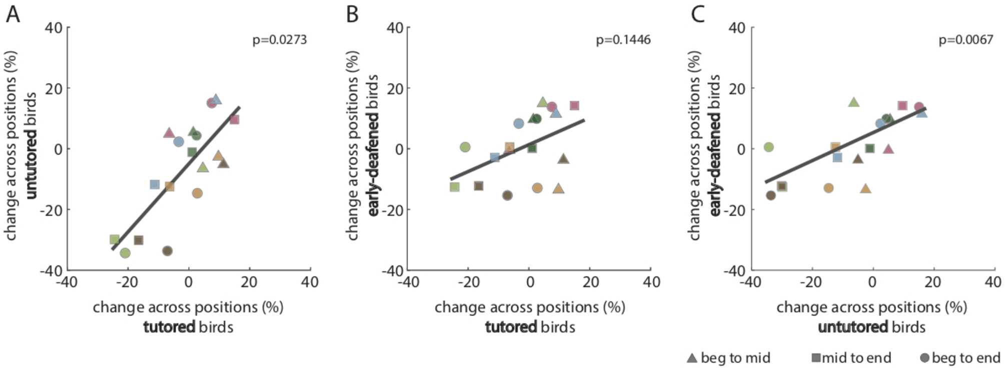

Figure 3

201x73mm (150 x 150 DPI)

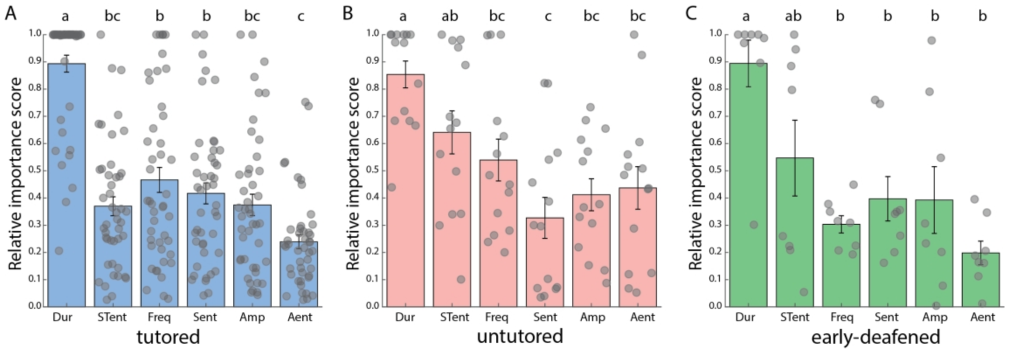
A
a
bc
b
b
bc
c
Relative importance score
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Dur
STent
Freq
Sent
Amp
Aent
tutored
B
a
ab
bc
c
bc
bc
Relative importance score
Dur
STent
Freq
Sent
Amp
Aent
untutored
C
a
ab
b
b
b
b
Relative importance score
Dur
STent
Freq
Sent
Amp
Aent
early-deafened

Figure 4

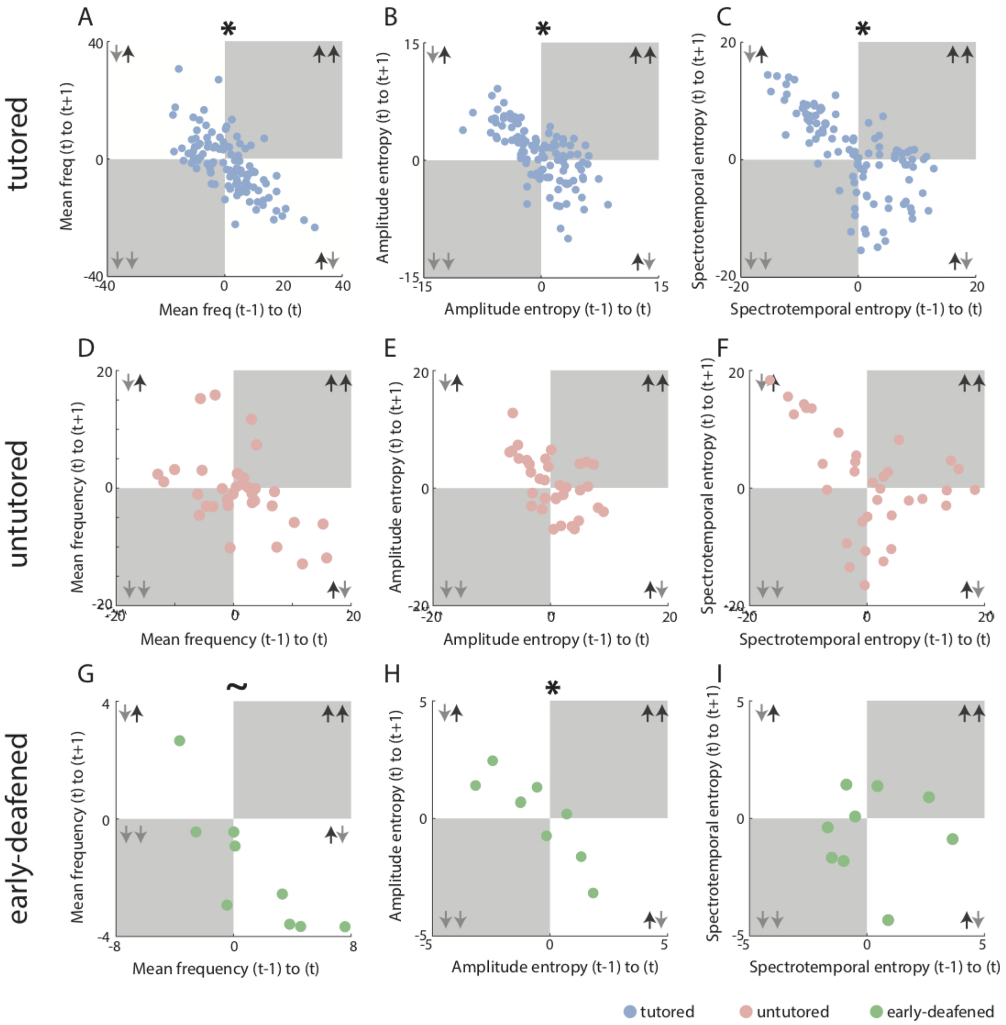

Figure 5

184x187mm (150 x 150 DPI)