Distribution Agreement

Yujingai Shi March 11, 2025

Investigating the Syntactical and Acoustic Control of Repeated Syllables In the Bengalese Finch

by

Yujingai Shi

Dr. Sam Sober
Advisor

Dr. Daniel Dilks
Co-Advisor

Psychology

Dr. Sam Sober

Advisor

Dr. Daniel Dilks

Co-Advisor

Dr. Hillary Rodman

Committee Member

Dr. Kristen Frenzel

Committee Member

2025

Investigating the Syntactical and Acoustic Control of Repeated Syllables In the Bengalese Finch

By

Yujingai Shi

Dr. Sam Sober
Advisor

Dr. Daniel Dilks
Co-Advisor

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Psychology

2025

Abstract

Investigating the Syntactical and Acoustic Control of Repeated Syllables In the Bengalese Finch

By Yujingai Shi

Understanding how animals regulate motor variability is important to uncovering the mechanisms of behavioral flexibility. To achieve external goals, animals must regulate their motor behaviors in a defined sequence of well-learned movements, thereby demonstrating a hierarchical structure of motor control. This principle also applies to birdsong, which exhibits variability at both syntactical (macro-level) and acoustic (micro-level) scales. However, it remains unclear whether such variability reflects structured regulation or arises from random processes. To address this gap, we investigate whether repeat syllables in Bengalese finch song as a form of learned vocal motor behavior would reflect structured regulations or random processes, or a combination of both. Using a dataset of spontaneous, well-learned vocalizations from an adult Bengalese finch, we analyzed sequential patterns and acoustic features of repeat syllables.

After demonstrating that repeat length probabilities deviated from both uniform and Markovian models, we applied a model fitting process; and the distribution was best modeled by a shifted log-normal distribution. This pattern suggests that repeat length is governed by multiplicative, feedback-sensitive mechanisms rather than random processes. Further analysis showed that a predictable set of three-syllable motifs preceded and followed repeat sequences regardless of length, indicating that repeats are probably embedded within larger structures.

Acoustically, we examined five features across repeat iterations: fundamental frequency (pitch), spectral entropy (noisiness), Wiener entropy (spectral flatness), syllable amplitude, and syllable duration. All these features individually showed a possible lack of directed acoustic drift. Instead, we observed a gradual drift in the first 4 syllables of a repeat sequences regardless of eventual length in global representations of all these acoustics in a lower-dimensional space.

In conclusion, these findings show that repeat syllables in birdsong possibly reflect a mix of pre-planned hierarchical control and drift either under neural control or due to biomechanical factors. Further works could use these salient behavioral features that reflect non-random control to explore how the brain organizes neural signals to drive variation. This work offers ideas into the mechanisms by which learned motor behaviors maintain stability and flexibility.

Investigating the Syntactical and Acoustic Control of Repeated Syllables In the Bengalese Finch

By

Yujingai Shi

Dr. Sam Sober
Advisor

Dr. Daniel Dilks
Co-Advisor

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Psychology

2025

Acknowledgements

Jonathan Vordzorgbe – Thank you for being the best mentor during my three and a half years in the lab. We started from knowing little things on projects together, and you taught me so many things on science and even beyond science.

Dr. Sam Sober – Thank you for your continuous support throughout my college journey! You are the best supervisor.

Dr. Amanda Jacob – Thank you for helping me along the way of forming my thesis!

Dr. Sober, Dr. Dilks, Dr. Rodman, Dr. Frenzel – Thank you for being on my committee and your continuous support and advice.

Ms. Ying Zhang and Dr. Bin Shi— Thank you, Mom and Dad, for bringing me to this world. So that I can have dreams, and I can have the courage to pursue my dreams.

Yufei Du – Thank you, my best friend.

Table of Contents

Investigating the Syntactical and Acoustic Control of Repeated Syllables In the Bengalese Finch

Introduction

The ability to flexibly generate complex motor behaviors is a significant challenge that several organisms must overcome; and they do this by generating neural signals that drive precisely learned behaviors that contain both stereotyped and variable features. Understanding the basis of such flexibility in motor output is essential for uncovering how organisms learn, adapt, and interact with their surroundings (Braun et al., 2009). One important domain for exploring flexibility in motor output is vocal communication. In both humans and other animals, vocalizations can be produced to convey meaningful social information, where the sequential and acoustic features of vocal gestures are precisely coordinated for functional significance (Taylor et al., 2022). Birdsong offers a compelling model for studying complex motor control since songbirds learn their vocalizations in a manner that resembles human speech learning: both involve a naïve organism relying on constant sensory reinforcement from a proficient social member; and both of the motor outputs are characterized by short, discrete units that are strung in variable sequences (Brainard & Doupe, 2013). Songbirds produce learned vocalizations as motor sequences that are readily quantifiable, rendering it a viable model to study the behavioral and neurological features of motor behaviors (Brainard & Doupe, 2013). Among songbirds, Bengalese finches are well-suited for investigating motor variability because they produce a rich, flexible vocal repertoire in a goal-directed manner (Kawaji et al., 2024). These vocalizations are apparently organized in a hierarchical structure, from individually distinct syllables with defined acoustic parameters such as pitch and amplitude to syntactical features that govern the sequence of these individual syllables in a larger song-bout. Interestingly, the Bengalese finch song contains a diversity of sequence features, with syllables arranged in repetitions or around sequence nodes, where distinct sequences can converge or diverge using a common syllable. This thesis aims to investigate the syntactical and acoustic structure of repeated syllables in Bengalese finch song to better understand how sequence and acoustic features reflect principles of motor control. By focusing on vocal repetition, a behavior that inherently involves both sequential and acoustic organization of the same syllable within the same song-bout, we aim to uncover the underlying dynamic organizations that support behavioral flexibility and potential hierarchical control.

Motor Control as a Significant Factor for Behavioral Complexity

Motor control enables organisms to execute a wide range of goal-directed behaviors, from simple actions to intricate, learned actions such as playing basketball, speaking, or singing. These learned skilled behaviors rely on the brain's capacity to coordinate movements in a flexible and adaptive manner due to various ethological demands. Therefore, during the process of executing such skilled behaviors, variability in motor output plays a critical role; it allows for exploration, adaptation, and fine-tuning of movement patterns based on environmental and contextual cues (Dhawale et al., 2017). For instance, the ability to "chunk" actions into discrete units may facilitate more efficient execution of various behaviors (Wu et al., 2023b). Therefore, understanding how these 'chunks' are both represented individually and sequenced within actions could shed further light on how the underlying neural code is organized.

Birdsong as a Model to Study Motor Control and Variability

Birdsongs, the vocalizations with social information produced by male songbirds, provide a powerful model system to study motor control and variability. Similar to human speech, birdsong is a learned vocal motor behavior that depends on early sensorimotor experience (Brainard & Doupe, 2002). During early development, songbirds memorize a "tutor" bird's songs as templates and refine and practice their learned song through auditory feedback (Brenowitz et al., 1997). After this “sensorimotor stage,” in their adulthood, songbirds are skilled vocalizers capable of maintaining both stereotyped and variable segments in their songs (Warren et al., 2012; Wohlgemuth et al., 2010). The consistent and stereotyped patterns in birdsong help maintain reliable social communications, whereas the ability to modify both syllable sequences and acoustics to add variability in response to contextual cues enables adaptive behaviors (Veit et al., 2021; Lipkind et al., 2017). Such interaction between stereotypy and variability reflects a functional feature of the motor system: balancing stability with variability.

The Bengalese finch, a type of songbird, is suited for investigating motor control and vocal variability. These birds sing spontaneously. For instance, they sing unprompted and produce a large number of iterations of vocalizations daily. Their songs are composed of discrete syllables and sequences that could be segmented and quantified into syntactical (e.g., transition probability) and acoustic (e.g., pitch) features (Seki et al., 2008; Sober et al., 2008). An adult bird's motor control system regulates “crystallized,” or well-learned stable birdsong to fall within a specified range of acoustic parameters; and the sequencing and phonology are connected, suggesting precise neural control (Wohlgemuth et al., 2010). Moreover, birdsong exhibits hierarchical variability: they not only exhibit variable sequential transitions but also display acoustic fluctuations in features like pitch, entropy, amplitude, and duration (Sober et al., 2008; Sakata & Brainard, 2008). This hierarchical structure provides an important model for studying how motor systems flexibly organize behavior.

In addition, the songbird brain contains discrete, well-defined nuclei that are specialized for song production and learning, forming pathways responsible for motor execution and vocal plasticity (Brainard & Doupe, 2002). This organized neural architecture enables consistent patterns of vocalizations, making it possible to observe how specific behaviors, such as sequence transitions and acoustic patterns, reflect underlying motor control processes. In addition, this makes it more accessible for in vivo neural recording than in humans. Therefore, it is now essential to carefully characterize the structure of song behavior, providing clear reference points for interpreting the underlying neural activity.

Learning Variability in Birdsong is Important for Understanding Motor Control

Studying variability in birdsong offers valuable insights into how motor behaviors are learned, refined, and flexibly controlled. While stereotypy in adult birdsong enables consistent communication, small fluctuations in both sequence and acoustic features persist in crystallized Bengalese finch songs. Rather than being purely random behaviors, these subtle variations are thought to reflect either adaptive or exploratory behaviors, supporting the idea that variability plays a key role in maintaining and regulating skilled motor output (Tumer & Brainard, 2007).

Using Repeat Syllables to Study Variability

Repeat syllables, characterized by showing multiple consecutive renditions of the same syllable, provide an ideal unit to investigate this question (James et al., 2021). Since the same syllable is produced repeatedly in a range of syntactic contexts, repeat sequences provide a controlled yet flexible structure for observing how vocal behavior varies within and across time. Their repetitive structure enables the analysis of both short-scale dynamics, such as individual acoustic modulation; and long-scale structure, such as global trends in repeat length distributions. Therefore, such capacity to reflect both higher-order sequence organization and real-time acoustic modulations makes repeats a strong candidate for studying how variability is expressed and regulated in learned vocal behavior. By focusing on repeat syllables, this study aims to determine whether the variability is random or reflects an underlying organization, suggesting hierarchical motor control processes.

Study Aims

The overarching goal of this study is to examine whether repeat syllables in adult Bengalese finch song exhibit structured variability in both sequence and acoustic domains. Although adult birdsong is a crystallized behavior, it displays subtle variability in sequential and acoustic layers (Wohlgemuth et al., 2010). This study investigates whether the parameters of this behavior show a pattern implied by random processes, or if they follow a distribution driven by a highly regulated process. Identifying whether or not there is a pattern to this behavioral variability is crucial in identifying what motor gestures may be governed by highly regulated neural signals in future studies.

The **central hypothesis** of this study is that variability in sequence and acoustics does not follow a random distribution but rather reflects an underlying organizational structure evidenced by non-random distribution patterns. The reason motivating this hypothesis is related to birdsong structures. To be specific, adult birds maintain flexibility in both the sequence of syllables and the acoustic parameters of repeated vocal elements. This suggests that even in stereotyped song, there is space for behavioral adaptation and reinforced motor exploration. While variability can arise from multiple sources, such as random noise in the motor system, consistent biased patterns of variation are more likely to reflect real-time organization. Because birdsong is a learned behavior that depends on defined neural circuits, the presence of structured variability would suggest precise motor control mechanisms embedded within the song system. The null hypothesis, in contrast, argues that the observed variability reflects a stochastic process with no systematic control to bias it in a specific direction.

To evaluate this, the study is investigating two domains: sequential features and acoustic features. In the domain of sequence, the aim is to determine first whether the number of syllable repetition, specifically “repeat length”, is governed by a structured organization or by randomness. One possible null model is a uniform distribution, where all repeat lengths are equally probable. Another possibility is a Markovian process, where each instance of repetition continues with a fixed probability less than one, independent of history which leads to an exponential decay in the probability of longer repeats. In both cases, the implication is that the bird exerts minimal real-time control over repeat length, and possibly initiates and terminates them stochastically. In contrast, if repeat length distributions differ from these null models, it would suggest real-time and/or history-dependent organization. In addition, the second test is to determine if repeat structure is influenced by preceding or succeeding syllables. If it is, this may suggest a global organization, supporting the idea that repeats are governed by a discernible syntactic framework. These two models were chosen as representative examples of random, memoryless processes, serving as minimal-control baselines.

To study acoustic structures, this study aims to investigate whether acoustic features such as pitch, spectral entropy, Wiener entropy, amplitude, and duration remain consistent across repeat iterations or whether they exhibit systematic drift during each repeat sequence. The rationale behind this test is that increased variability in some social contexts (e.g., undirected vs. directed song) suggests the ability to control this output in a meaningful way (Fortkord & Veit, 2025). Repeats might represent this type of control within a song sequence as opposed to globally over multiple bouts. If there is a drift, it may suggest either top-down neural control, or gradual changes in vocal output due to physical limitations or fatigue such as muscular tension or respiratory effort, or a combination of both (Day et al., 2008). If there is no drift, it could suggest top-down control to ensure no systemic drift, with variability existing due to physical limitations, but not significant enough to cause a systematic and discernable drift. The nature of the drift and its correlation with the repeat dynamics may show the interplay between macro- and micro-level control. This paints a better picture of the precise neural control as well as what role physical factors may play.

The null hypothesis suggests that each syllable repetition is acoustically drawn from the same distribution, resulting in no significant relationship between repeat iteration number and acoustic features. Such a pattern would suggest either a lack of internal modulation and reflect fixed representations of each syllable, or tight regulation of each rendition of a syllable within a defined range regardless of the potential biochemical challenges of rapidly repeating the same motor gesture. However, if acoustic features change predictably across iterations, such as gradual increases, this suggests either dynamic control of motor output or currently uncharacterized downstream biomechanical influences during the course of a repeat sequence.

Therefore, investigating sequential and acoustic features of repeat syllables would provide a comprehensive framework for testing whether repeat variability in birdsong is random or organized. By identifying structured patterns sequence and acoustic features, this study aims to reveal more underlying processes of motor control in learned behavior.

Methods

Data Collection and Preprocessing

The dataset used in this study consists of a series of song recordings from a single adult Bengalese finch (Bird ID: bl136yw14), collected across 11 days (Fig. 1A). We collected ‘undirected songs’ – which are sung alone (not at a female) and tend to be more variable in acoustics. To prepare the data for analysis, we first segmented continuous vocalizations into discrete song bouts based on inter-syllable gap durations. Following criteria set from previous studies (Zhang et al., 2017), we set a threshold of 350 milliseconds (ms) as the maximum silent gap between vocalized syllables as part of the same song sequence. Gaps longer than this value were treated as indicators of the end of a song bout.

To extract song structure and sequence patterns, we constructed a matrix of syllable-to-syllable transitions and generates summarized descriptive statistics for all detected syllables. The transition matrix enabled us to identify syllables with high self-transition probabilities (e.g., "h → h"). We then selected those syllables most likely to be repeated consecutively as candidate 'repeat syllables' for the main analysis in the next step.

Main Analysis: Identifying and Extracting Repeat Syllables

We extracted and quantified all instances of repeat bouts within the identified song sequences. In addition, for each of the extracted recordings, we counted the repeat length (i.e., the number of consecutive repetitions of the repeat syllable within that bout) and generated distributions of these measures to assess any patterns in repeat statistics (Fig. 1A). Several studies of songbird repeat statistics have demonstrated a non-random, normally distributed pattern, so we expected to find as such in this bird (Fujimoto & Hasegawa, 2023). We decided to further explore the motifs flanking these repeat sequences, as well as possible patterns in the sequence dynamics around repeats of varying lengths.

Motifs flanking repeat sequences up to 3 syllables long were analyzed to assess history-dependence as well as transition dynamics around repeat sequences, providing insights into exploring syntactical control. We performed these analyses for both long and short repeat sequences to clarify their overall structure by testing whether repeat lengths could be predicted by preceding motifs, and/or if they further predicted the following syllables.

Main Analysis: Sequence Analysis of Repeat Syllables

To evaluate whether the structure of repeat syllables reflects non-random variability, we examined the distribution of repeat lengths across all extracted recording days (11 days, 1915 song-bouts). Although adult birds produce learned, crystallized song, they still show variability in how many times a syllable is repeated in succession.

To demonstrate this, as well as further characterize the behavior of this bird, we first established two null hypotheses in this analysis. The first null model assumes a uniform distribution of repeat lengths, where every possible repeat length (e.g., 2, 3, or 4 repeats in succession) is equally probable (Fig. 2A). This model represents a scenario in which the bird selects a repeat length randomly, with little to no evidence of a structured syntax that would result from precise neural control. The second null model tests a Markovian process, where the probability of repeating a syllable depends only on a fixed self-transition probability, without history dependence (Fig. 2B). Under this null model, repeat sequences decay exponentially in probability as length increases. For example, for a coin with a probability of 0.5 to land on heads, the probability of flipping 10 heads consecutively is significantly lower than for 5 heads in a row, and more for flipping 15 heads. This null model represents a system where individual syllables have a fixed probability to transition to all other syllables, including itself. Under this system, this fixed probability to repeat the syllable ensures that longer repeat bouts are less probable and occur spontaneously without any structured organization to bias the total length in a direction.

Next, to test the empirical distribution of repeat lengths against these null models, we generated null datasets. For the uniform model, we created a null uniform distribution that took the most commonly produced repeat lengths of the bird (3-9 syllables long; empirical range 2-11, but extremes excluded due to low count) and set their probabilities as equal to each other. We tested this null distribution with the empirical distribution of repeat lengths. We performed a Chi-Square Goodness-of-Fit Test to assess whether the empirical repeat length distribution differed significantly from this model that represents random distribution. For the Markovian model, we simulated a first-order Markov process; we simulated repeat lengths using a fixed self-transition probability from the empirical data and compared this synthetic distribution to the empirical data. A p-value below 0.05 in either test was taken as evidence against the null hypothesis, supporting the hypothesis that repeat lengths are not generated randomly. Because the test against a Markov model was done with a simulated distribution that could change slightly on each simulation, we bootstrapped this analysis to determine that the findings were not due to chance.

After demonstrating a deviation from the uniform and Markov models as previously shown, we further sought to determine the most accurate model to fit our empirical data. We first compared several candidate distributions to assess which best captured the shape of the empirical repeat length data. Given that repeat lengths are discrete, we tested multiple commonly used distributions for modeling count data: the normal distribution, the gamma distribution, and the negative binomial distribution. Under assumptions of a normal distribution, the repeat lengths would center around a mean, with few outliers, suggesting a strong preference in the bird for a specific length. However, this would fail to capture heavy skew in the behavior, which is also possible, especially for non-random behaviors that are not merely stochastic variations centered around a mean. For this reason, we also tested other models (gamma, negative binomial, log-normal), which are more robust in representing the potential skew or bias in the data.

Based on the previous assumption that repeat generation reflects a combination of internal and possibly external influences, combined with the fact that the repeat distribution only contains non-negative values, we hypothesized that a log-normal distribution would be an appropriate candidate model for the data. To address issues of non-zero lower bounds, we introduced a shift parameter to stabilize the log-transformation and applied Maximum Likelihood Estimation (MLE) to fit the log-normal model, estimating its mean (μ) and standard deviation (σ). We used the Akaike Information Criterion (AIC) to evaluate model fit, with lower AIC values indicating better performance. Additionally, we tested a frequency-weighted version of the log-normal model to account for the disproportionate occurrence of certain repeat lengths. This approach allowed us to determine whether a more flexible distribution could better characterize the data than the rigid null models; and also determine which parameters of this model best describe aspects of the behavior. We performed this analysis on individual days as well as on the pooled dataset across all 11 days of recording and observed similar results as would be expected in a well-learned behavior.

Finally, we examined whether the structure of repeat sequences was influenced by their surrounding song context. Specifically, we examined the motifs that occurred immediately before and after repeat bouts. <u>Our hypothesis was that certain unique motifs stereotypically and systematically precede or follow repeat sequences of particular lengths.</u> To test this, we extracted 3 syllables before and after each repeat bout and categorized these pre- and post-repeat motifs by the length of the repeat. We then computed the frequency and probability distributions of these motifs across different repeat lengths.

To determine whether contextual patterns varied systematically with repeat length, we compared motif distributions across repeat groups. If specific motifs reliably occur before or after certain repeat lengths, that would suggest that repeats and their lengths are dynamically controlled online and may be predicted with some history-dependence. The presence of structured transitions into and out of repeat bouts would provide strong evidence for short-term dynamic control of lengths as opposed to a fixed and hierarchical structure. However, the lack of any unique flanking motifs would suggest that repeats are embedded within larger global structures and modulated only internally.

Main Analysis: Acoustic Analysis of Repeat Syllables

To learn more about the structure of repeat syllables, we then examined whether the acoustic structure of repeat syllables varies systematically across repetitions. We extracted and analyzed five fundamental acoustic features: fundamental frequency (pitch), spectral entropy, Wiener entropy, amplitude, and duration. These features were selected since they captured both the harmonic content, such as frequency, and the temporal dynamics within individual birdsong syllables. To be specific, pitch reflects the fundamental frequency of a syllable, spectral entropy measures the complexity or noisiness of the spectral distribution, and Wiener entropy quantifies the diversity of power across frequency (Moorman et al., 2021). Amplitude represents the strength of the vocal signal, and duration calculates the time between the onset and offset of the syllable.

First, to investigate acoustic features’ changes across repeat iterations (e.g., the first, second, and third syllables within a sequence of repeated syllables), we grouped all repeat sequences. In order to ensure sufficient samples, we restricted our analyses to repeat sequences between 4 and 9 syllables long. However, inclusion of the rarer, longer, and shorter repeat bouts had no significant effect on our findings. We then computed summary statistics across iterations: for each of the five features, we calculated the mean and the standard deviation for the first four syllables to ensure that all repeat bouts were analyzed regardless of how long they lasted beyond the 4 first syllables. Along with the means, we calculated the standard error of the mean (SEM) to assess variability across various sequences.

Then, to quantify whether there are specific trends that exist in either the mean or standard deviation of acoustic features across iterations and if there is a correlation between these and the progression of the repeat. We computed the Spearman correlation coefficients (ρ) and p-values; then, we fitted linear regression lines. This approach enabled us to assess whether the average and variance (spread) of these acoustic measures increased, decreased, or remained stable across individual syllables within a single repeat bout.

Finally, we performed a Principal Component Analysis (PCA) to assess the overall structure of syllables in the acoustic features. For this, we first created a matrix containing all syllables within repeat sequences, with rows corresponding to syllables and columns representing the five acoustic features. We conducted PCA on the matrix to reduce dimensionality and plotted the first three components to identify potential clusters using their average Euclidean distance. This was done with the goal of testing the hypothesis that the syllables with the same index in a repeat sequence are arranged in distinct clusters. In particular, we were testing for evidence of a systemic drift which would be present in distinct clusters in PC-space of syllables that correlate with repeat iteration.

Together, these analyses were accomplished using a constructed MATLAB code pipeline which allowed extraction and alignment of acoustic data by repeat structure. Across all the analyses, this pipeline enabled us to examine both the sequential organization and acoustic dynamics of repeats by generating repeat indices, aligning acoustic features by iteration number, and applying statistical techniques.

Results

Sequential Features of Repeat Syllables

To investigate whether repeat length distributions in Bengalese finch song exhibit structure beyond stochastic processes, we first analyzed the frequency distribution of repeat length for a commonly produced target syllable with high repeat probability (78%), “h.” The distribution revealed a positively skewed, heavy-tailed pattern (Fig. 1D). In this distribution, most of the repeat sequences (N = 1532) clustered between 3 and 6 syllables long.

Next, when we tested whether the empirical distribution of repeat lengths deviated from two null models, one assuming equal probability of all repeat lengths (uniform) and the other based on fixed syllable self-transition probabilities (first-order Markov), we found significant differences in both cases. Compared to the uniform model, the empirical distribution showed significant deviation from the uniform distribution (Chi-square test, $p < .001$), suggesting that repeat length selection is not uniform (Fig. 2A). Compared to the Markovian model, the empirical distribution also revealed significant deviation, demonstrating that repeat length selection is not a Markovian distribution (Fig. 2B).

Next, when we fitted several models to the empirical repeat length data to assess which distribution best accounted for the observed distribution from the empirical data, we found the best-fit model to the empirical distribution. Among the candidate models discussed in the above section (Normal, Gamma, Negative Binomial, and Log-Normal), the Log-Normal model with a frequency-weighted fit provided the best match with the empirical distribution, as indicated by the lowest AIC score ($AIC = 4.23$) showing the best fit (Fig. 2C). This supports the idea that repeat length distribution is governed by structured variability and likely influenced by several multiplicative pre-planning factors rather than random output (Guerrero & Garcia-Baños, 2020).

Finally, we examined the sequential context in a broader range to examine the potential global organization of repeat structures; we analyzed the motif sequences immediately preceding and following repeat sequences. For the syllable "h," one of the pre-repeat 3-syllable motifs, "efg" showed a dominant structure across all repeat lengths (Fig. 3A), and one of the pre-repeat 3-syllable motifs, "yjj" similarly exhibited a dominant structure consistently following all repeat sequences, regardless of length (Fig. 3B). While these motifs were most common, they were not exclusive: alternative motifs occurred in a small proportion of cases (~5–10%). These minority patterns may reflect flexible transitions or context-dependent variability. To capture this broader distribution, we further demonstrated the breadth of this behavior using heatmaps (Fig. 3C), showing that specific motifs are reliably associated with particular repeat sequences "efg" and "yjj." These findings are based on one high-repeat syllable ("h"), and future analyses will be necessary to assess whether similar patterns hold for other syllables, such as "d."

These results of sequence analyses suggest that repeat syllables are embedded within structured sequential contexts instead of governed by random processes; and that their length distributions reflect planning and higher-order organization rather than dynamic probabilistic transitions.

Acoustic Features of Repeat Syllables

We plotted the overall distribution of five fundamental acoustic features: fundamental frequency, spectral entropy, Wiener entropy, amplitude, and duration across all repeat syllables (Fig. 1C). These distributions provided a reference to understand both the central tendency and the variability in the dataset. First, fundamental frequency (pitch) showed a unimodal distribution centered around ~1580 Hz, with a slight right skew. This pattern suggests a dominant pitch range used by this bird for this syllable, with occasional higher-pitch renditions. Second, spectral entropy showed a normal distribution, centered around 0.8. Third, Wiener entropy also showed a roughly symmetric distribution centered around -10, showing relatively consistent frequency dispersion across syllables. Fourth, amplitude showed a strong positive skew, with most syllables having relatively low intensity and fewer syllables produced at higher amplitudes. Finally, duration was slightly skewed to the right, with most syllables ranging between 45 and 55 milliseconds, indicating a relatively consistent temporal structure with some variability in length. These distributions illustrate that while the bird's vocalizations are generally consistent, there exists notable acoustic variability which provides a foundation for investigating how these properties evolve across repeat iterations.

Then, when we evaluated whether acoustic properties change systematically over the course of repetition, we analyzed how the mean value of each acoustic feature evolved across successive iterations within a repeat sequence (Fig. 4A). For each feature, we calculated the mean and standard error of the mean (SEM) at each syllable position (from 1 to 4) within the repeat. In general, all of the five features—pitch, spectral entropy, Wiener entropy, amplitude, and duration—showed no significant changes in mean or variance with iteration number, indicating no systemic acoustic drift or change in within individual bouts. Previous studies have shown a decrease in variance of acoustics under global social contexts (Sakata & Brainard, 2008). Therefore, it is interesting to see a lack of modulation in these individual measures within repeats during a social context where songbirds are known to exhibit the most acoustic variation (undirected song). These findings overall suggest that acoustic features do not individually undergo gradual modulation across syllable repetitions.

Next, to assess whether the range of acoustic variability changed across repeat iterations, we analyzed the standard deviation of each feature at each syllable position (Fig. 4B). This analysis allowed us to examine not only how average acoustic properties (mean values, Fig. 4A), but also to see if there were potential top-down control of this variability or biomechanical noise that increased or decreased during the repeat sequence. Overall, trends in standard deviation were not pronounced. First, fundamental frequency exhibited a moderately increasing trend in variability, though this correlation was not statistically significant. Amplitude showed a similar moderate upward trend, suggesting a possible but non-definitive increase in loudness variability over time. These results suggest that the within-iteration variability remained relatively stable, indicating that repeat-specific modulations may be driven more by biochemical noise.

In addition, to further explore how acoustic features evolve across repeat iterations, we conducted a Principal Component Analysis (PCA) on the five extracted acoustic features (Fig. 5A). PCA reduces the dimensionality of the dataset while preserving the most informative component, allowing us to assess whether different iterations of a repeated syllable cluster distinctly in acoustic feature space. The PCA clustering visualization (Fig. 5A) showed that the first 4 syllables of every repeat sequence (>= 4 syllables long) exhibited some distinct clusters in this lower-dimensional space. Further inspection of the clustering analysis and visualization reveals that there is a clear global drift in the acoustics as the repeat continues (Fig. 5B). This may be shown acoustically in the increase in duration and possibly amplitude (not significant but trending), although this might be more apparent with increased data sizes. We emphasize that this finding does not identify neural control directly, but it does demonstrate the presence of a drift in acoustic variability that may co-occur with macro-level syntactic control. We also performed a Chi-sqaure test to show that the global acoustic drift is statistically significant ($p < .001$). A further hypothesis that we could not test due to sample size issues is the rate of this drift along successively longer repeat sequences.

Discussion

Summary of Findings

This study investigated the sequential and acoustic structure of repeat syllables in Bengalese finch song to determine whether such variability is random or regulated by underlying structures. The results demonstrated several key findings. First, the empirical distribution of repeat lengths deviated significantly from both a uniform model and a first-order Markov model, as shown previously, indicating that simple probabilistic mechanisms could not account for the observed variability. Instead, a shifted log-normal distribution provided the best fit to the empirical data, suggesting the presence of structured and perhaps hierarchical motor planning towards a slightly skewed distribution. Along with apparently being embedded within larger song structures, as evidenced by stereotyped flanking motifs regardless of repeat length, it is reasonable to conclude that this bird is maintaining relatively fixed structures that encompass a wide range of outputs by internally modulating lengths of repeats. This would suggest that the bird is producing a pre-planned set of song sequences with varying repeat lengths and possibly does little online modulation of these sequences outside of the actual repeat. While the bird maintains a relatively fixed structure, the wide range of possible repeat lengths within that structure points to structured variability: a form of controlled flexibility rather than randomness. This distinction is especially important in the context of social communication, as female songbirds are known to evaluate male songs based on both the stability of syntax and the range or precision of acoustic features. Prior work has shown that directed songs, produced in courtship, often exhibit reduced variability and more stereotyped structure compared to undirected songs (Chen et al., 2017). Thus, understanding how repeat variability is structured may help explain how males balance consistency with flexibility in different social contexts.

In addition, when investigating acoustic structure through PCA, we found that although individual acoustic features did not show systematic changes, the PCA analysis across the five features suggests a clear and global systematic acoustic drift through the repeat iterations. In future studies, this feature may shed light on how the dynamics at sequence 'nodes' (syllables that serve as convergence or divergence points for different sequences) are controlled. The clustering pattern suggests an evolving structure, possibly shaped by top-down modulation.

Interpretation of Findings in the Context of Motor Control

Our results support the view that motor output in birdsong is modulated by dynamic and hierarchical control processes, rather than by random generation. The finding that repeat lengths follow a log-normal distribution is particularly informative. Log-normal distributions frequently emerge from multiplicative processes, where variability accumulates across sequential steps—for example, in human reaction times or animal foraging durations—indicating that recurrent behavior may be influenced by internal feedback mechanisms or resource limitations within the brain control system (Guerrero & Garcia-Baños, 2020). This pattern is distinct from the exponential decay predicted by Markov models.

These findings are consistent with the broader idea that variability in motor output is not merely random noise but may reflects flexible and adaptive control. Prior work has emphasized the functional role of motor variability in exploration and adaptation (Dhawale et al., 2017). Here, the structured variation in repeat length supports this idea; rather than generating identical vocal sequences, Bengalese finches appear to fine-tune their motor output in real time to meet their expectations, potentially allowing for context-dependent adjustments.

Deafening studies in songbirds (Sakata & Brainard, 2006) further support this idea by demonstrating that auditory feedback is critical for maintaining the structure and precision of learned song—once removed, variability increases, suggesting that fine control is learned and actively maintained.

By maintaining a variable set of parameters for the same gesture, we believe it is important to determine which regions of this distribution appear to be correlated with other salient features in the song (e.g., initiation of a repeat, increased tempo, etc.). This could be explored by aligning acoustic parameter distributions to specific song events and comparing subsets (e.g., early vs. late repeats) using correlation analyses or conditional probability estimates. These findings would point more to a learned, planned, and possibly ethologically relevant behavior instead of being influenced strongly by noise.

Additionally, social context plays a key role in modulating this variability. Prior work by has shown that variability is actively suppressed in directed song (e.g., when courting a female), whereas it is enhanced in undirected song (Sakata & Brainard, 2009), highlighting that variability itself may be under behavioral control depending on communicative goals.

Interplay between Sequence and Acoustic Structure in Song

The consistent, non-random structure of repeat sequences in the song as well as the relatively fixed, but wide range of acoustics suggests that these birds produce songs from a stable repertoire. They appear to have a set of songs-bouts they pick from, even if they do so in a non-random manner. Outside of the global central tendency, we did not observe any significant predictive power, based on prior history at any point within the song to determine how long a repeat would be sustained once it was initiated. Neither did the bird exhibit any significant change in further vocalizations after repeats of varying lengths were produced. This suggests a very weak influence of biomechanical factors at the macro-level (sequence). Conversely, the presence of a non-random acoustic drift in global parameters suggests a significant neural drive and/or predictable biomechanical influences that may depend on the tuning and physiology of the vocal system's machinery.

Implications for Neural Control of Birdsong

These behavioral findings have important implications for understanding the neural circuits involved in birdsong. Specifically, they raise questions about how hierarchical structures are maintained and modulated in the brain. The basal ganglia nucleus Area X, known to mediate song variability and reinforcement learning, is a strong candidate for controlling repeat behavior (Fujimoto & Hasegawa, 2023; Sakata & Brainard, 2008). Previous studies have shown that basal ganglia support real-time adaptation and could play a role in the modulation of variability (Fee & Goldberg, 2011). However, studies of the avian motor cortex have shown the presence of highly predictable, stereotyped outputs that drive song-bouts with a high degree of sequence consistency (Duffy et al., 2019). It is therefore important to identify clear evidence of salient behavioral control if we are to hope to understand the underlying neural code. The structured variability in both sequential and acoustic features suggests that distinct neural states may accompany the progression of a repeat sequence, potentially reflecting changes in planning or motor execution over time. Additionally, understanding this neural mechanism will help clarify the role of biomechanical factors as well, and allow us to determine what aspects of the behavior are under tight control or truly due to random noise.

Moreover, parallels between birdsong and human speech are important to explore in an in-depth manner. Like songbirds, humans exhibit dynamic modulation of repetitions in speech; and both can exhibit stuttering behavior under a range of basal ganglia perturbations (Fujimoto & Hasegawa, 2023). While it is clear that human speech exhibits syntactical use of repeats in some contexts, it is not yet clear how and even why birds employ these. In order to understand this, it is important to demonstrate what aspects of this behavior appear to be more structured. The presence of variability in both systems suggests that similar principles of motor control and hierarchical planning may apply. Thus, understanding how songbirds manage repetition in vocal behavior could provide insights into speech motor control and its neural mechanisms in humans.

Limitations and Future Directions

While the present findings are compelling, several limitations should be discussed. Most importantly, the study was conducted on data from a single bird recorded across multiple days. We performed the same analysis on a different bird and discovered similar patterns in both sequence stability and acoustic variation (data not shown). However, these other analyses were limited by smaller sample sizes or noisy acoustic data. Although these birds produced a rich and diverse set of vocalizations, it is essential to replicate these findings across multiple individuals to assess generalizability. Differences in individual repertoire size, motivation, or age could influence repeat behavior, and a broader sample would clarify the extent of variability.

In addition, our analysis focused on one repeat syllable, although the bird actually maintained two other syllables as repeats. For one of them, ‘d,’ the bird actually produced more iterations of this repeat; however, it was a highly noisy variable with acoustics difficult to reliably quantify, along with multiple apparently truncated syllables. More powerful techniques to better segment and quantify these syllables are crucial to fully address the central question of vocal control Other repeat syllables were too infrequent and could not be justified for analysis due to the need for a large sample size as well as reducing the effects due to the possibly imprecise nature of our quantification analyses.

In conclusion, this study reveals that repeat syllables in Bengalese finch song are regulated by structured variability in sequential domains and acoustic domains. Rather than being random outputs, repeat sequences reflect dynamic, feedback-sensitive control mechanisms that are embedded in a broader song syntax. These findings highlight the Bengalese finch as a powerful model for understanding how the brain generates flexible and organized motor behaviors, offering insights relevant not only to birdsong but also to human speech and general motor control systems.

Tables and Figures

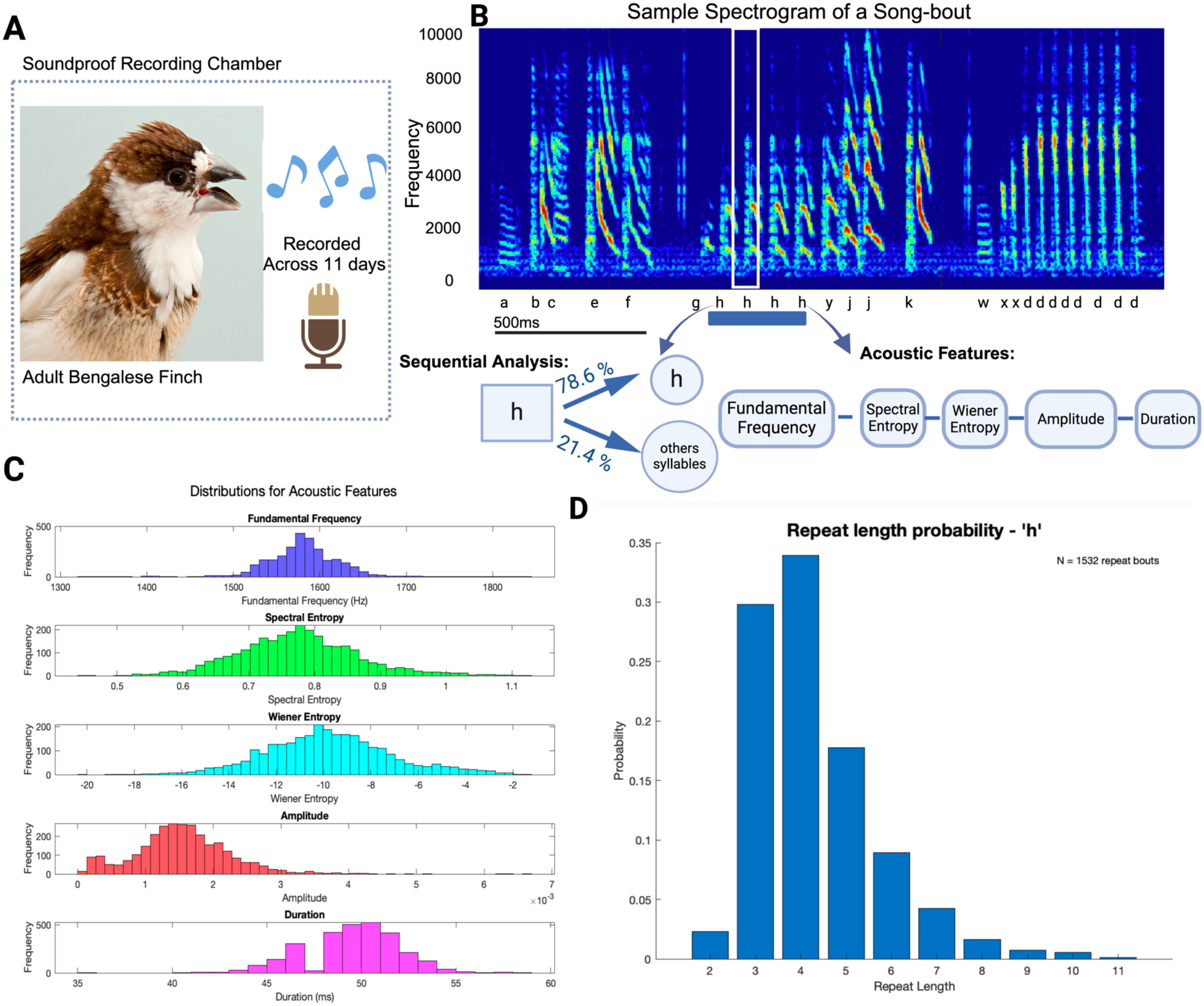

Figure 1. *Experimental paradigm and acoustic analysis of repeat syllables in Bengalese finch song*
(A) Crystallized adult Bengalese finch songs were recorded and used as the primary dataset for analysis. Songs were segmented into syllables based on acoustic features and syllable identity. (B) Example spectrogram of a song bout from one bird (bl136yw14). Syllables are annotated beneath the spectrogram. This study focuses on analyzing repeat syllables, such as “h”, which occur consecutively within bouts. (C) Five acoustic features were extracted for each syllable: Fundamental Frequency, Spectral entropy, Wiener entropy, Amplitude, and Duration. These features were used to examine acoustic drift and structural regularities across repeated syllables. (D) Probability distribution of repeat lengths (number of consecutive syllable repetitions) for the syllable “h”, based on acoustic identity. The distribution shows a heavy-tailed pattern, with most repeat bouts clustering between 3 and 6 repetitions (N = 1532 repeat bouts).

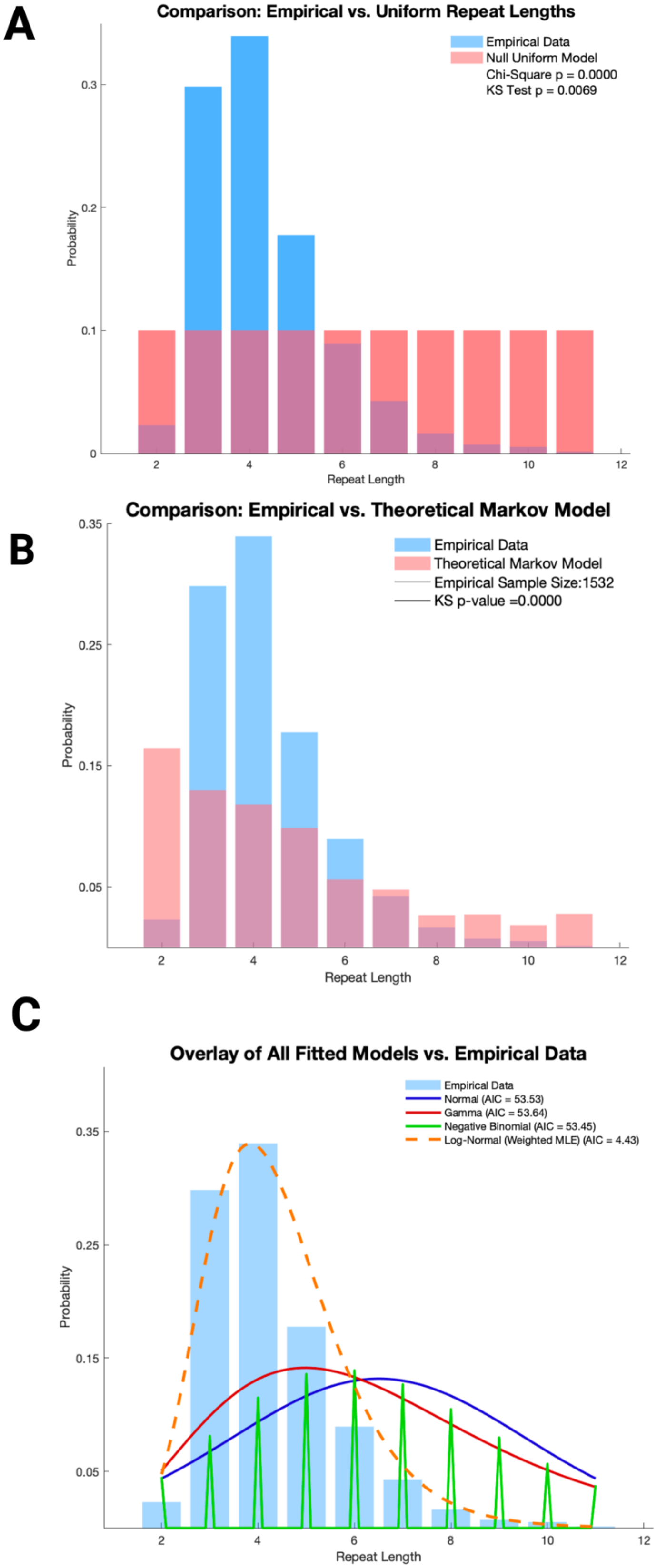

Figure 2. *Statistical Analysis of Sequential Features*
(A) Comparison between the observed repeat length distribution and a null uniform model. The empirical data deviates significantly from uniformity (Chi-square and KS test, $p < 0.05$), indicating that repeat lengths are not random. (B)Empirical distribution compared to a null Markovian model generated from syllable transition probabilities. The two distributions are significantly different (KS test, $p < 0.05$), suggesting the repeat structure is not explained solely by simple transition rules. *(*C) Model Fitting to Empirical Data: Overlay of four fitted models—Normal, Gamma, Negative Binomial, and Log-Normal (weighted MLE)—on the empirical distribution. The Log-Normal model provides the best fit, supported by the lowest AIC score.

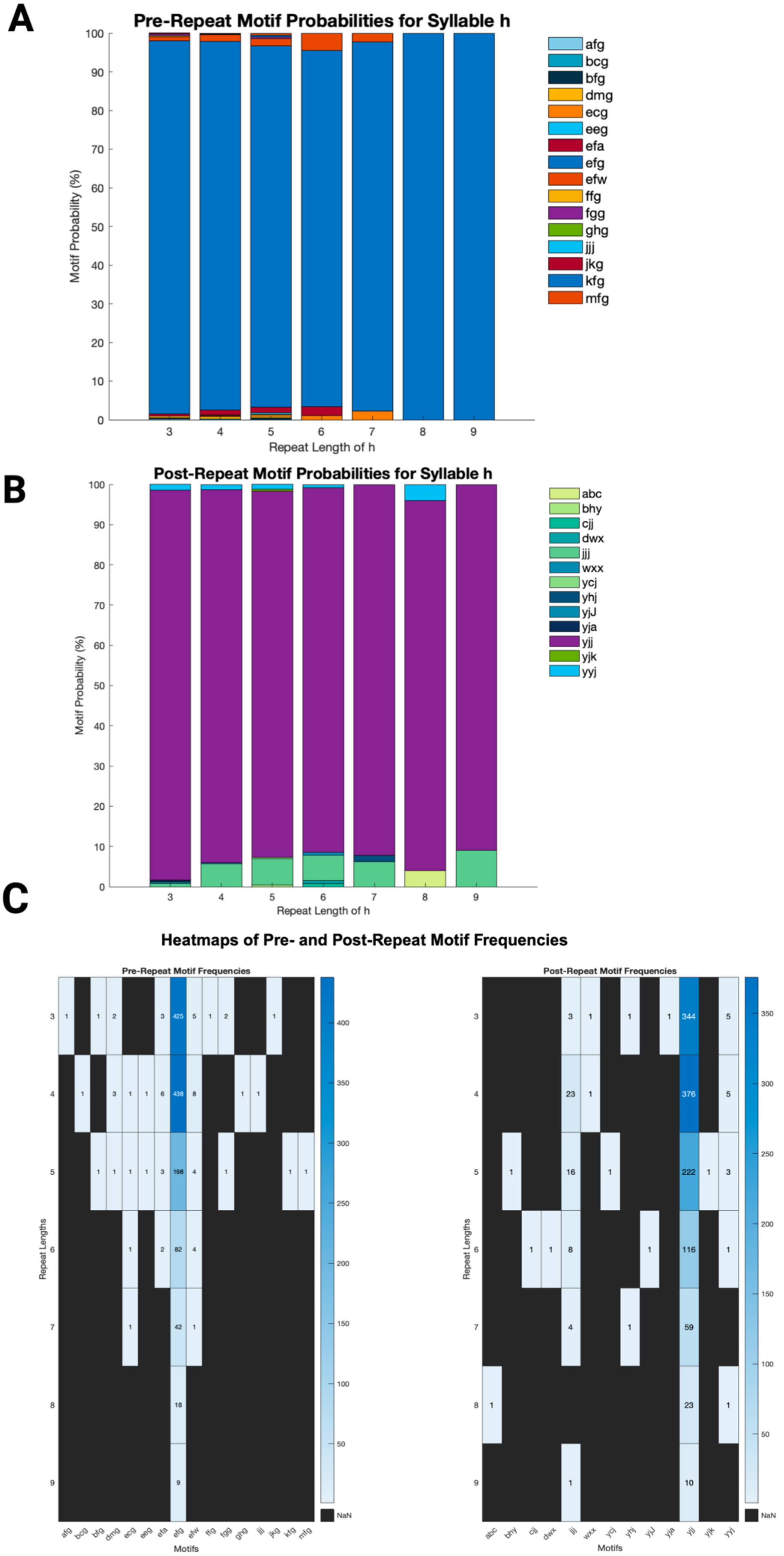

Figure 3. *Contextual Analysis of Sequential Features*
(A) Pre-Repeat Motif Probabilities: The probability distribution of motifs immediately preceding the repeat syllable 'h'. One motif occurs consistently before repeats across all repeat lengths, suggesting a structured transition into repetition. (B) Post-Repeat Motif Probabilities: Motif distribution following the repeat. Similar to (A), post-repeat transitions appear constrained, with one dominant motif observed after the repeats. *(C)* Heatmap of Pre- and Post-Repeat Motif Frequencies:Matrix visualization of motif occurrences before and after repeat bouts. Each cell shows the frequency count for a motif given a specific repeat length, highlighting structured sequential dependencies.

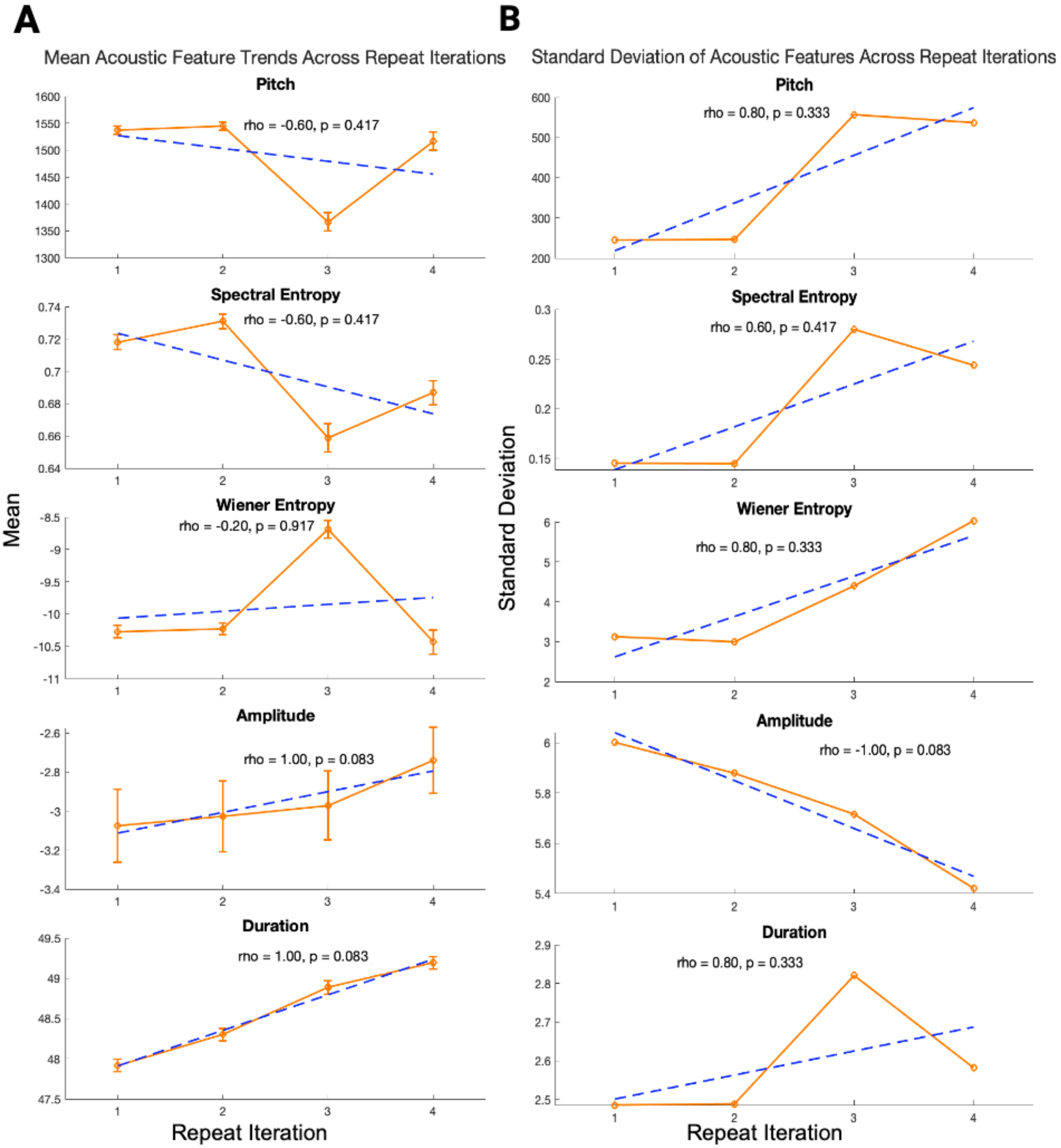

Figure 4. *Statistical Analysis of Acoustic Features*
(A) Mean acoustic feature trends across repeat iterations shows the mean ± SEM values of five key acoustic features (Pitch, Spectral Entropy, Wiener Entropy, Amplitude, and Duration) across successive syllable positions within a repeat sequence. Each subplot includes a linear fit line and Spearman correlation coefficient), highlighting how feature values evolve over the course of repeated vocalizations. (B) Standard deviation of acoustic features across repeat iterations panel visualizes the standard deviation of each feature across repeat iterations. It reflects how within-iteration variability shifts over the course of repetition and includes correlation metrics for assessing systematic trends.

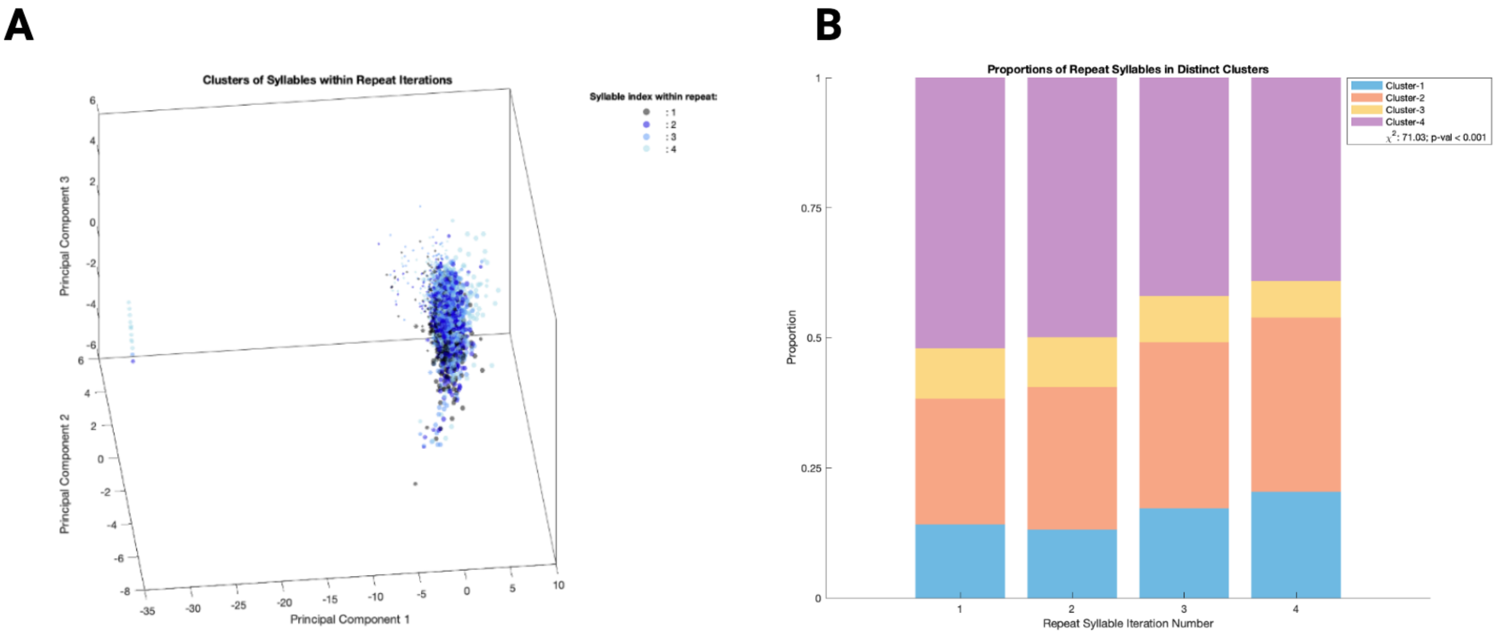

Figure 5. *Principal Component Analysis (PCA) on Repeat Acoustics*
(A) *Principal Component Analysis (PCA) of acoustic feature panel displays a 3D scatter plot of PCA scores for all repeated syllables, colored by their position within the repeat sequences.*
(B) Proportions of repeat syllables in PCA clusters. Bar plot showing the proportions of repeat syllables assigned to each of the four clusters (Cluster 1–4), grouped by their iteration number within the repeat. Cluster composition shifts across iterations, suggesting changes in the global acoustic profile of repeated syllables over time.

References

Brainard, M. S., & Doupe, A. J. (2002). What songbirds teach us about learning. *Nature*, *417*(6886), 351–358. https://doi.org/10.1038/417351a

Brainard, M. S., & Doupe, A. J. (2013). Translating birdsong: songbirds as a model for basic and applied medical research. *Annual review of neuroscience, 36*, 489–517. https://doi.org/10.1146/annurev-neuro-060909-152826

Braun, D. A., Aertsen, A., Wolpert, D. M., & Mehring, C. (2009). Learning optimal adaptation strategies in unpredictable motor tasks. *Journal of Neuroscience*, *29*(20), 6472–6478. https://doi.org/10.1523/jneurosci.3075-08.2009

Brenowitz, E. A., Margoliash, D., & Nordeen, K. W. (1997). An introduction to birdsong and the avian song system. *Journal of Neurobiology*, *33*(5), 495–500. https://doi.org/10.1002/(sici)1097-4695(19971105)33:5

Chen, Y., Clark, O., & Woolley, S. C. (2017). Courtship song preferences in female zebra finches are shaped by developmental auditory experience. *Proceedings of the Royal Society B Biological Sciences*, *284*(1855), 20170054. https://doi.org/10.1098/rspb.2017.0054

Day, N. F., Kinnischtzke, A. K., Adam, M., & Nick, T. A. (2008). Top-Down regulation of plasticity in the birdsong system: “Premotor” activity in the nucleus HVC predicts song variability better than it predicts song features. *Journal of Neurophysiology*, *100*(5), 2956–2965. https://doi.org/10.1152/jn.90501.2008

Dhawale, A. K., Smith, M. A., & Ölveczky, B. P. (2017). The Role of Variability in Motor Learning. *Annual review of neuroscience, 40*, 479–498. https://doi.org/10.1146/annurev-neuro-072116-031548

Duffy, A., Abe, E., Perkel, D. J., & Fairhall, A. L. (2019). Variation in sequence dynamics improves maintenance of stereotyped behavior in an example from bird song. *Proceedings of the National Academy of Sciences*, *116*(19), 9592–9597. https://doi.org/10.1073/pnas.1815910116

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

Fortkord, L., & Veit, L. (2025). Social context affects sequence modification learning in birdsong. *Frontiers in Psychology*, *16*. https://doi.org/10.3389/fpsyg.2025.1488762

Fujimoto, H., & Hasegawa, T. (2023). Reversible inhibition of the basal ganglia prolongs repetitive vocalization but only weakly affects sequencing at branch points in songbirds. *Cerebral Cortex Communications*, *4*(3). https://doi.org/10.1093/texcom/tgad016

Guerrero, F. G., & Garcia-Baños, A. (2020). Multiplicative processes as a source of fat-tail distributions. *Heliyon*, *6*(7), e04266. https://doi.org/10.1016/j.heliyon.2020.e04266

James, L. S., Mori, C., Wada, K., & Sakata, J. T. (2021). Phylogeny and mechanisms of shared hierarchical patterns in birdsong. *Current Biology*, *31*(13), 2796-2808.e9. https://doi.org/10.1016/j.cub.2021.04.015

Jin, D. Z., & Kozhevnikov, A. A. (2011). A compact statistical model of the song syntax in Bengalese Finch. *PLoS Computational Biology*, *7*(3), e1001108. https://doi.org/10.1371/journal.pcbi.1001108

Kawaji, T., Fujibayashi, M., & Abe, K. (2024). Goal-directed and flexible modulation of syllable sequence within birdsong. *Nature Communications*, *15*(1). https://doi.org/10.1038/s41467-024-47824-1

Lipkind, D., Zai, A. T., Hanuschkin, A., Marcus, G. F., Tchernichovski, O., & Hahnloser, R. H. R. (2017). Songbirds work around computational complexity by learning song vocabulary independently of sequence. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-01436-0

Mason G. J. (2010). Species differences in responses to captivity: stress, welfare and the comparative method. *Trends in ecology & evolution, 25*(12), 713–721. https://doi.org/10.1016/j.tree.2010.08.011

Medina, I., & Francis, C. D. (2012). Environmental variability and acoustic signals: a multi-level approach in songbirds. *Biology Letters*, *8*(6), 928–931. https://doi.org/10.1098/rsbl.2012.0522

Moorman, S., Ahn, J., & Kao, M. H. (2021). Plasticity of stereotyped birdsong driven by chronic manipulation of cortical-basal ganglia activity. *Current Biology*, *31*(12), 2619-2632.e4. https://doi.org/10.1016/j.cub.2021.04.030

Sakata, J. T., & Brainard, M. S. (2006). Real-Time contributions of auditory feedback to avian vocal motor control. *Journal of Neuroscience*, *26*(38), 9619–9628. https://doi.org/10.1523/jneurosci.2027-06.2006

Sakata, J. T., & Brainard, M. S. (2009). Social context rapidly modulates the influence of auditory feedback on avian vocal motor control. *Journal of Neurophysiology*, *102*(4), 2485–2497. https://doi.org/10.1152/jn.00340.2009

Seki, Y., Suzuki, K., Takahasi, M., & Okanoya, K. (2008). Song Motor control organizes acoustic patterns on two levels in Bengalese finches (Lonchura striata var. domestica). *Journal of Comparative Physiology A*, *194*(6), 533–543. https://doi.org/10.1007/s00359-008-0328-0

Sober, S. J., Wohlgemuth, M. J., & Brainard, M. S. (2008). Central contributions to acoustic variation in birdsong. *Journal of Neuroscience*, *28*(41), 10370–10379. https://doi.org/10.1523/jneurosci.2448-08.2008

Taylor, D., Clay, Z., Dahl, C. D., Zuberbühler, K., Davila-Ross, M., & Dezecache, G. (2022). Vocal functional flexibility: what it is and why it matters. *Animal Behaviour*, *186*, 93–100. https://doi.org/10.1016/j.anbehav.2022.01.015

Tumer, E. C., & Brainard, M. S. (2007). Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. *Nature*, *450*(7173), 1240–1244. https://doi.org/10.1038/nature06390

Veit, L., Tian, L. Y., Hernandez, C. J. M., & Brainard, M. S. (2021). Songbirds can learn flexible contextual control over syllable sequencing. *eLife*, *10*. https://doi.org/10.7554/elife.61610

Warren, T. L., Charlesworth, J. D., Tumer, E. C., & Brainard, M. S. (2012). Variable sequencing is actively maintained in a well learned motor skill. *Journal of Neuroscience*, *32*(44), 15414–15425. https://doi.org/10.1523/jneurosci.1254-12.2012

Wittenbach, J. D., Bouchard, K. E., Brainard, M. S., & Jin, D. Z. (2015). An adapting auditory-motor feedback loop can contribute to generating vocal repetition. *PLoS Computational Biology*, *11*(10), e1004471. https://doi.org/10.1371/journal.pcbi.1004471

Wohlgemuth, M. J., Sober, S. J., & Brainard, M. S. (2010). Linked control of syllable sequence and phonology in birdsong. *Journal of Neuroscience*, *30*(39), 12936–12949. https://doi.org/10.1523/jneurosci.2690-10.2010

Wu, S., Éltető, N., Dasgupta, I., & Schulz, E. (2023b). Chunking as a rational solution to the speed–accuracy trade-off in a serial reaction time task. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-31500-3

Zhang, Y. S., Wittenbach, J. D., Jin, D. Z., & Kozhevnikov, A. A. (2017). Temperature manipulation in Songbird brain implicates the premotor nucleus HVC in Birdsong syntax. *Journal of Neuroscience*, *37*(10), 2600–2611. https://doi.org/10.1523/jneurosci.1827-16.2017