Within-species differences in vocal production learning in a songbird are associated with differences in flexible rhythm pattern perception

Andrew A. Rouse[1], Aniruddh D. Patel[1, 2], Samantha Wainapel[3], and Mimi H. Kao[4, 5]

[1]Department of Psychology, Tufts University; [2]Program in Brain, Mind and Consciousness, Canadian Institute for Advanced Research; [3]Tufts University; [4]Department of Biology, Tufts University; [5]Neuroscience Graduate Program, Tufts University.

*Corresponding author: Andrew A. Rouse

Email: andrew.rouse@tufts.edu

Media summary

When we recognize a song that's played faster or slower than usual, we are detecting timing *relations* between sounds, not just absolute durations. This ability relies on auditory-motor interactions in the brain. Auditory-motor processing is more developed in animals that learn their vocalizations so we hypothesize that such animals are advantaged in recognizing a rhythmic pattern at different rates. We test this idea in a songbird where only males learn to sing, and find males are slightly better than females at rhythmic pattern recognition. Thus, songbirds are ideal for investigating individual differences in rhythm pattern perception and underlying brain mechanisms.

Abstract

Humans readily recognize a familiar rhythmic pattern, such as isochrony (equal timing between events) across a wide range of rates. This ability reflects a facility with perceiving the relative timing of events, not just absolute interval durations. Several lines of evidence suggest that this ability is supported by precise temporal predictions that arise from forebrain auditory-motor interactions. We have shown previously that male zebra finches, which possess specialized auditory-motor networks and communicate with rhythmically patterned sequences, share our ability to recognize isochrony independent of rate. To test the hypothesis that flexible rhythm pattern perception is linked to vocal learning, we ask whether female zebra finches, which do not learn to sing, can also recognize global temporal patterns. We find that non-singing females can flexibly recognize isochrony but perform slightly worse than males on average. These findings are consistent with recent work showing that while females have reduced forebrain song regions, the overall network connectivity of vocal premotor regions is similar to that in males and supports predictions of upcoming events. Comparative studies of male and female songbirds thus offer an opportunity to study how individual differences in auditory-motor connectivity influence perception of relative timing, a hallmark of human music perception.

1. Introduction

The ability to recognize auditory rhythms is important for many species [1–3], but the underlying neural mechanisms are only beginning to be understood. One area of progress is recognition of communication signals based on tempo. Female field crickets, for example, are attracted by male calling songs composed of trains of short sound pulses when the tempo is ~ 30 syllables/sec. This selectivity is hard-wired and mediated by a small network of interneurons to processes instantaneous pulse rate [4]. While this preference is genetically fixed, in other animals, experience can sculpt neural responses to behaviorally-salient call rates. For example, in the mouse auditory cortex, excitatory cells are innately sensitive to the most common pup distress call rate (~5 syllables/s), but their tuning can broaden to a wider range of rates following co-housing with pups producing calls across a range of rates [5].

Much less is known about how the brain recognizes rhythmic *patterns* independently of rate. While humans can encode and remember the rate of auditory sequences [6], we also readily recognize rhythmic patterns across a broad range of rates [7]. In music, one of the most widespread temporal patterns is isochrony, or equal timing between events [8] and the ability to detect and predict periodicity is central to the positive effects of music-based therapies on a variety of neurological disorders, including normalizing gait in Parkinson's disease [9,10]. The ability to recognize a rhythmic pattern whether it is played fast or slow is present in infants [11] and is based on recognition of the relative timing of events more than on their absolute durations. In humans, there is growing evidence that the neural mechanisms underlying detection of relative timing are distinct from those involved in encoding absolute timing [12–14]. In addition, neuroimaging studies have shown that both auditory *and* motor regions are active when people listen to rhythms, even in the absence of overt movement. Responses in several motor regions are greater when the stimulus has a strong, periodic pulse, or "beat" [15,16] and transient manipulation of auditory-motor connections using transcranial magnetic stimulation disrupts beat perception without affecting single-interval timing [17].

Based on such findings, we and others have suggested that perception of temporal regularity (independent of rate) depends on the interaction of motor and auditory regions: the motor planning system uses information from the auditory system to make predictions about the timing of upcoming events, and communicates these predictions back to auditory regions via reciprocal connections [18,19]. Such predictions could support detection of rhythmic patterns independent of tempo because the relative duration of adjacent intervals remains the same across different rates (e.g., 1:1 for an isochronous pattern).

Given that vocal learning species often communicate using rhythmically patterned sequences [20,21] and have evolved specialized motor planning regions that are reciprocally connected to auditory forebrain regions, we have hypothesized that vocal learners are advantaged in flexible auditory rhythm pattern perception [22]. Consistent with this hypothesis, zebra finches and starlings (vocal learning songbirds) trained to discriminate isochronous from arrhythmic sound sequences robustly generalize this ability to stimuli at novel tempi, including rates distant from the training tempi [23,24]. This ability, also seen in humans [7], demonstrates a facility with recognizing a rhythm based on global temporal patterns, since absolute durations of intervals differ markedly at distant tempi. These findings contrast with similar research conducted with vocal non-learning species. For example, pigeons can learn to discriminate sound sequences based on tempo but cannot learn to discriminate isochronous from arrhythmic stimuli [25]. Rats can be trained to discriminate isochronous from arrhythmic rhythms, but show weak generalization when tested at novel tempi, suggesting a strong reliance on absolute timing for rhythm perception [26].

Here, we further test the hypothesis that differences in vocal learning abilities correlate with differences in flexible rhythmic pattern perception by taking advantage of the sexual dimorphism in zebra finches [27]. In this species only males learn to imitate song, and the neural circuitry subserving vocal learning is greatly reduced in females. Thus we predicted that male zebra finches would exhibit faster learning rates for discriminating isochronous versus arrhythmic stimuli and/or a greater degree of generalization in recognizing these categories at novel tempi. By using the same apparatus, stimuli, and methods, we can meaningfully compare the flexibility of male vs. female rhythmic pattern perception in this species.

Several prior findings, however, support the opposite prediction - either no sex differences in rhythmic pattern perception or better rhythmic pattern perception in female zebra finches. One reason females may not be disadvantaged in our task is that they analyze male song when choosing a mate, and a previous meta-analysis found that female zebra finches are faster than males at learning to discriminate spectrotemporally complex auditory stimuli [28]. In addition, while the volume of vocal motor regions that subserve song performance is substantially greater in male zebra finches, a recent anatomical study found male-typical patterns of connectivity in the vocal premotor region in females, including minimal sex differences in afferent auditory and other inputs [29]. Moreover, a recent study of antiphonal calling suggests that female zebra finches may perform better than males on rhythmic pattern processing [30]. In that study, both male and female zebra finches could predict the timing of calls of a rhythmically calling vocal partner, allowing them to adjust the timing of their own answers to avoid overlap. This ability to predictively adjust call timing was enhanced in females and was disrupted by lesions of vocal motor forebrain regions, suggesting that brain regions associated with singing in males may subserve auditory perception and/or timing abilities in female zebra finches.

2. Methods and analysis

(a) Subjects

Subjects were 24 experimentally naïve female zebra finches from the Tufts University breeding colony (mean age=72 ± 8 (SD) days post-hatch (dph) at the start of training; range=61-87 dph). Data from 14 age-matched male zebra finches were collected previously [22] (mean age=74 ± 11 dph at start of

training, range=62-106 dph, spanning the end stages of sensorimotor learning). All procedures were approved by the Tufts University Institutional Animal Care and Use Committee.

(b) Auditory Stimuli

All stimuli used in this experiment were the same as those described previously [22]. Briefly, we used sequences of natural sounds from unfamiliar male conspecifics: either an introductory element that is typically repeated at the start of a song, or a short harmonic stack commonly sung by zebra finches [31] (see Fig. S1). For each sound, an isochronous sequence with equal time intervals between event onsets and an arrhythmic sequence with a unique temporal pattern (but the same mean inter-onset interval (IOI)) were generated at two base tempi: 120 ms and 180 ms IOI (Fig. 1A). These tempi were chosen based on the average syllable rate in zebra finch song (~7-9 syllables per second, or 111-142 ms IOI [31]). The number of elements, overall duration, amplitude, and spectral profile were matched across a pair of isochronous and arrhythmic stimuli with a given sound and tempo, and the average duration of the sequences was 2.32 ± 0.1 (SD) s (20 intervals and 13 intervals for the two training tempi, respectively). For the arrhythmic stimuli, the minimum gap between consecutive elements was 0 ms, and the maximum possible IOI was 1.5 times the base IOI. (see Fig. S2 for the coefficient of variation of IOIs for each arrhythmic stimulus). An additional pair of probe stimuli (one isochronous, one arrhythmic) was generated at a tempo (144 ms IOI) 20% faster than the slower training stimulus (180 ms IOI) and 20% slower than the faster training stimulus (120 ms IOI).

For “rule training” (see below), an isochronous/arrhythmic pair was generated with a novel harmonic stack (sound D, see Fig. S1) at tempi ranging from 75 ms to 275 ms IOI in 1 ms steps (i.e., at 201 rates). Each arrhythmic stimulus had a temporally unique pattern. For a subset of these stimuli (n=15/201), the amplitude was inadvertently higher as described previously [22].

(c) Auditory Operant Training Procedure

Training and probe testing used a go/interrupt paradigm as described previously [22]. In these experiments, birds had access to a trial switch and a response switch, with a water reward spout in the response switch. Female birds were mildly water restricted and worked for water rewards (~5-10 µl/drop), routinely performing ~510 trials/day on average. Briefly, pecking the trial switch triggered playback of a stimulus: either rewarded (S+, 50% chance) or unrewarded (S-, 50% chance). For all rhythm discrimination experiments the S+ stimuli were isochronous patterns and the S- stimuli were arrhythmic. Trial and response switches and their associated lights were activated 500 ms after stimulus onset, and pecking either switch after this time halted playback. “Hits” were correct pecks of the response switch during S+ trials, resulting in a water reward. “False alarms” were pecks of the response switch on S- trials, and resulted in lights out to act as a mild aversive. The duration of the lights-out period (2–25 s) was adjusted to control for response bias; the period was lengthened if a bird tended to respond to every stimulus [32]. If neither switch was pecked within 5 s of trial onset, the trial would end, and was marked as ‘no response’. No response to the S+ stimulus was considered a “miss”, while no response to the S- stimulus was counted as a “correct rejection”. During this 5 s window, the bird could also peck the trial switch again to “interrupt” the current trial with no lights-out consequences, which was counted as a ‘miss’ or ‘correct rejection’ depending on whether a S+ or S- stimulus had been presented. The use of the trial switch to interrupt trials varied widely among birds, and therefore was not analyzed further in this study. Regardless of response, birds had to wait 100 ms after the stimulus stopped playing before a new trial could be initiated.

Shaping and Performance Criteria. To learn the go/interrupt procedure, birds were first trained to distinguish between two unfamiliar conspecific songs (~2.4 s long, “shaping” phase), one acting as the S+ (rewarded or ‘go’) stimulus, and the other as the S- (unrewarded or ‘no-go’) stimulus. Lights-out

punishment was not implemented until a bird demonstrated reliable usage of the switches by performing ≥ 100 S+ trials. In all shaping and training phases, the criterion for advancement to the next phase was ≥ 60% hits, ≥ 60% correct rejections, and ≥ 75% overall correct for two of three consecutive days. Five females did not complete the shaping process within 30 days and were removed from the study and excluded from further analysis.

Training: *rhythm discrimination*. Once a bird reached criterion performance on discriminating the shaping stimuli, she was trained to discriminate between isochronous and arrhythmic stimuli (n=13 birds). As described previously, each bird was trained using multiple sound types and multiple stimulus rates. In the first phase of training, each bird learned to discriminate two isochronous stimuli (120 ms and 180 ms IOI) from two arrhythmic stimuli (matched for each mean IOI). Once performance criterion was reached, a new set of stimuli at the same tempi but with a novel sound element (and a novel irregular temporal pattern at each tempo) was introduced (see Fig. 1B). Subjects learned sounds in one of two orders: one group of females ('ABC'; n=6) was trained to discriminate sound A stimuli, followed by sound B and then sound C; a second group ('CAB'; n=7) started with sound C, followed by sounds A and B. One female 'ABC' bird was presented with stimuli at 3 additional tempi (137.5 ms, 150 ms, and 157.5 ms IOI), but those trials were not reinforced and were excluded from analysis.

Probe Testing/Generalization. To assess whether birds could generalize the isochronous vs. arrhythmic classification to a novel (untrained) tempo, they were tested with probe stimuli at 144 ms IOI, 20% slower /faster than the training stimuli (120 ms and 180 ms IOI). We previously showed that male zebra finches can discriminate between two stimuli that differ in tempo by 20% [22] and confirmed this in female zebra finches (see below and Fig. S3B). Probe sounds were introduced only after the bird had successfully completed two phases of training (e.g., either sounds A and B or sounds C and A; see Fig. 1B). Prior to probe testing, the reinforcement rate for training stimuli was reduced from 100% to 80% for at least two days in order to acclimate birds to an occasional lack of reinforcement. During probe testing, training stimuli were presented in 90% of trials, and probe stimuli were randomly interleaved in 10% of trials. Probe trials were never reinforced or punished. 10% of the interleaved training stimuli also were not reinforced or punished, so the lack of reinforcement was not unique to probe stimuli [22,33].

Rule Training. Following probe testing with all three sound types, subjects (n=5 birds) were presented with a new set of isochronous and arrhythmic stimuli using a novel sound (sound D). In this phase, stimulus tempi included every integer rate from 75 ms to 275 ms (201 total rates or 402 stimuli). Arrhythmic stimuli were again generated independently so that each rate had a novel, random irregular pattern. Trials were randomly drawn from the set with replacement, and all responses were rewarded or punished as during training. This larger stimulus set made it unlikely that subjects could memorize individual stimulus patterns.

Discrimination of other acoustic features. To determine whether the 20% difference between the training and probe stimuli could be detected by female zebra finches, we tested a separate cohort of females on tempo discrimination (n=6; 71 ± 9.1 (SD) dph at start of shaping). Following shaping, these birds were trained using isochronous sequences of sound A at two rates: 120 ms IOI and 144 ms IOI. The rewarded (S+) stimulus was 120 ms IOI for four females and 144 ms IOI for two females. These females were also tested on their ability to discriminate spectral features using frequency-shifted isochronous sequences of sound A: one shifted 3 semitones up, and the other shifted 6 semitones up ("Change pitch" in Audacity v 2.1.2). Four females were tested with the 6-semitone shift as the S+ stimulus, and two females were tested with the 3-semitone shift as the S+ stimulus.

(d) Data Analysis

All statistical tests, except for the binomial tests for training and generalization, were performed in R (v. 3.6.2) within RStudio (v. 1.2.5033). Binomial logistic regressions were performed with lme4 (glmer) statistical package for R.

Training and generalization testing. To quantify overall performance during training and probe testing, for each bird, the proportion of correct responses ((Hits + Correct Rejections) / Total number of trials) was computed for each stimulus pair (isochronous and arrhythmic patterns of a given sound at a particular tempo). For training phases, proportion correct was always computed based on the last 500 trials. For probe testing, performance was based on proportion correct during the 80 probe trials. The proportion of correct responses was compared to chance performance (p=0.5) with a binomial test using α=0.05/2 in the training conditions (2 tempi), α=0.05/3 for probe trials (3 tempi). A linear least-squares regression was used to examine the correlation between each subject's average performance on interleaved training stimuli and probe stimuli.

Performance of females (n=7) during probe trials was compared to that of male zebra finches (n=7) collected in a prior study [1]. Performance across all probe trials (n=240 probe trials, 80 probe trials per song element x 3 song elements) was analyzed with a binomial logistic regression using a generalized linear mixed model with sex, probe block number, and sex-by-probe block number interaction as fixed effects and subject as a random effect.

Rule training. As described previously [22], the first 1000 trials were analyzed for each bird (n=5) to minimize any potential effect of memorization. Trials were binned in 10 ms increments starting at 75 ms IOI, and the number of correct responses was analyzed with a binomial logistic regression using a generalized linear mixed model with tempo bin as a fixed effect and subject as a random effect. Performance in each bin was compared to performance in the 75-85 ms IOI bin in which each subject's performance fell to chance as the degree of temporal variation in inter-element intervals was severely limited by the duration of the sound element (see Fig. S2). To identify possible sex differences, we performed an additional analysis using data from a range of tempi (95-215 ms) in which both males and females performed significantly above chance ($p<0.005$; binomial test). The number of correct and incorrect trials were calculated per sex and analyzed for significant group differences with a $\chi 2$ test.

Discrimination of Other Acoustic Features. Performance on tempo or frequency discrimination was analyzed in the same manner as for rhythm discrimination: the proportion of correct responses in the last 500 trials was compared to chance performance using a binomial test with α=0.05.

Reaction Time. For each trial where the bird pecked a switch, the time between trial initiation and response selection was recorded. This value was averaged for all switch responses for the last 500 trials within a training phase. For the rhythm discrimination birds, only the first training block (prior to probe testing) was used. Mean RT values were compared between 14 birds that completed all rhythmic training phases, 9 birds that failed to complete rhythm discrimination training, and 10 birds (that learned the frequency discrimination (n=10; data from 6 females in this study and from 4 males collected as part of [22]) using a two-way ANOVA (sex, group, and sex-by-group interaction) followed by a Tukey multiple comparisons post-hoc test. Data met criteria for homoscedasticity and normal error distribution.

3. Results

(a) Rhythmic pattern training and generalization testing in female zebra finches

To test the ability of female songbirds to recognize a rhythmic pattern based on the relative timing of events, 13 female zebra finches were first trained on rhythm discrimination using a modified go/no-go paradigm with multiple training phases (Figs. 1A–B & S1; age: 61–87 dph on the first day of training). Seven out of 13 females learned to discriminate isochronous versus arrhythmic sequences in all three phases, each of which used a different zebra finch song element. The number of trials that it took for

females to reach criteria for successful discrimination (see Materials and Methods) was comparable to that of male zebra finches tested previously (Fig. 1C; data from male birds in [22]). Fig. 2A shows the time course of learning for a representative female zebra finch (*y7o97*). Initially, this female responded to both the isochronous and arrhythmic stimuli, but then gradually learned to withhold her response to the arrhythmic stimulus. Across the seven females, rhythm discrimination performance was ~81% accurate at the end of the first training phase (Fig. 2B, left column), and birds attained a comparable accuracy level by the end of each training phase (median = 80% proportion correct; see Fig. S3A for learning curves for each bird on the rhythm discrimination task).

After successful completion of two phases of rhythm discrimination with two different sound elements, females were then asked whether they could generalize the discrimination of isochronous versus arrhythmic stimuli at a novel tempo (Fig. 1B, 'probe testing'). Fig. 2B shows the performance on randomly interleaved training stimuli (*dark gray bars*; 90% of the trials*)* and 80 probe stimuli (*light gray bars;* 10% of the trials) for the seven females that successfully completed all three phases of rhythm discrimination training. Performance on probe stimuli was significantly above chance for 17 of 21 probe tests (n=3 probe tests/bird x 7 birds, $p < 0.0167$, binomial test with Bonferroni correction), indicating that female zebra finches can readily distinguish between the isochronous and arrhythmic stimuli at the new tempo (Fig. 2B).

(b) Sex differences in rhythm pattern perception

To test the hypothesis that the ability to flexibly perceive rhythmic patterns is linked to vocal learning, we directly compared performance of female zebra finches, which do not learn to imitate song, with that of vocal learning males (Fig. 3). Across all generalization tests, discrimination of isochronous versus arrhythmic stimuli at a novel tempo (probe stimuli) was significantly lower for female zebra finches compared to males (n=7 males and 7 females; 240 probe trials per bird; mixed-effects logistic regression, $p = 0.033$). The observed sex difference in the ability to generalize rhythm discrimination did not reflect differences in motivation to perform the task. On average, female zebra finches performed ~520 trials/day, compared to ~530 trials/day for males in our previous work. Across all subjects (n=14), performance on training stimuli (learned discrimination) was positively correlated with performance on probe stimuli (measure of generalization; Fig. S4; $p = .0375$).

As a second test of the ability to recognize temporal regularity independent of tempo, five females that completed rhythm discrimination training were tested with sequences of a new sound element (sound D, see Figs. S1-S2) across a wide range of tempi (75-275 ms IOI; "rule training"). On a given trial, one of 402 different stimuli was presented (equal probability of isochronous and arrhythmic sequence), reducing the likelihood that correct discrimination was based on memorization. Fig. 4 shows the average performance of all birds over the first 1000 trials, broken into 10 ms bins (~50 trials per bin). For female birds, performance was best between 95 and 215 ms IOI (tempi ~20% slower to ~25% faster than the original training range: darker bars with >=2 black asterisks, $p<0.005$; mixed effect logistic regression). Performance fell to chance at faster tempi (75-85 ms IOI), when the degree of temporal variation in inter-element interval durations was limited by the duration of the sound element (see Fig. S2), and at slower tempi (225-275 ms IOI). To directly compare performance to that of male zebra finches, the number of correct and incorrect trials within the 95-215 ms IOI tempo range (where both sexes performed well above chance) were grouped by sex. In this range, female performance was significantly worse than male performance ($\chi^2(1) = 13.919$, $p<0.001$).

(c) Reaction times for discrimination of acoustic features.

Across female birds who succeeded in rhythm discrimination training (n=7), the average reaction time at the end of the first training phase was ~1.70 s after stimulus onset, or ~75% of the duration of the stimulus train (Fig. 5). Data from previously tested males that succeeded in rhythm discrimination training

(n=7) shows a similar average reaction time (1.90 s), indicating that both males and females heard 10-15 intervals on average before responding (overall average reaction time of ~1.80 s). In contrast, the average reaction time for the females that did not reach the criterion for rhythm discrimination during training (n=6) was ~1.30 s, ~50% of the duration of each stimulus, similar to the average failed male (n=3) reaction time of ~1.18 s.

Shorter reaction times, however, do not necessarily indicate poor perceptual discrimination. In a separate cohort of birds that learned to discriminate isochronous sequences that differed in frequency by 3 semitones (i.e., a quarter of an octave; n = 6 females; average proportion correct responses in last 500 trials = 91%; see Fig. S3B and n = 4 males from [1]), the average reaction times for successful discrimination was ~1.24 s, consistent with the notion that unlike rhythm discrimination, discrimination based on spectral features does not require hearing long stretches of the sequence.

Comparison of reaction times between successful rhythm discrimination birds, unsuccessful rhythm discrimination birds, and successful frequency discrimination birds showed a significant effect of group (Fig. 5; two-away ANOVA, $F(2,27)$, $p = 0.001$). Tukey post-hoc tests showed that the reaction times of the successful rhythm birds were significantly higher (slower) than both the unsuccessful rhythm birds and the successful frequency birds. There was no effect of sex or a sex-by-group interaction. Taken together, these results suggest that success at rhythm discrimination may be related to how long birds listened to the stimuli before responding. Indeed, even among the birds that successfully generalized the discrimination of temporal regularity across a wide range of tempi, reaction time was longer for the tempi where the discrimination was more difficult (i.e., when the proportion of correct responses was closer to chance performance: 75-85 ms and > 235 ms; Fig. S5).

4. Discussion

To test the hypothesis that differences in the capacity for flexible rhythmic pattern perception correlate with differences in vocal learning, we investigated the ability of a sexually dimorphic songbird to recognize a fundamental rhythmic pattern common in music and vocal communication – isochrony [8,20,21]. In zebra finches and many other songbirds, only the males learn to sing and possess pronounced forebrain motor regions for vocal production learning [27]. Previously, we showed that male zebra finches can readily discriminate isochronous versus arrhythmic patterns independent of absolute time intervals [22], and we predicted that female zebra finches would perform less well when tested with the same rhythm discrimination and generalization tasks. Using a sequential training paradigm with multiple sound types and tempi, we found that about half of females tested (n=7/13) could learn to differentiate isochronous from arrhythmic patterns, and the learning rates for these females was similar to that of males (Fig. 1C). Females that completed rhythm discrimination training readily generalized the discrimination to stimuli at novel tempi (Fig. 2B). Correct discrimination of isochronous vs. arrhythmic stimuli was significantly above chance at tempi ranging from 20% slower to ~25% faster than the original training range (Fig. 4), indicating that female zebra finches, like males, can recognize a rhythmic pattern based on the relative timing of events rather than on absolute durations. Notably, in both sexes birds that succeeded in learning to recognize isochrony listened for longer before responding compared to birds who failed (Fig. 5, left two bars), consistent with attention to global temporal patterns.

While female zebra finches can learn to recognize isochrony independent of tempo, we found small but significant differences in performance between males and females. First, across all three tests of generalization at a novel tempo, males outperformed females (Fig. 3; average proportion of correct responses during probe trials: 75% vs. 69%). Second, males consistently slightly outperformed females when tested with stimuli across a broad range of tempi at which both sexes performed well above chance (Fig. 4; 95-215 ms IOI; average proportion correct: 76% vs 72%;). These differences did not reflect differences in motivation to perform the task; on average, males and females performed comparable numbers of trials/day. Finally, the proportion of females that successfully recognized isochrony based on

global temporal patterns (n=7/13) was lower than that of males tested previously (n=7/10), although more data would be needed to determine if this difference is reliable.

It is important to note that the sex difference we observed, while consistent with our hypothesis, is an *average* difference. Modest, but consistent, sex differences are well known in biology (e.g., on average, men are taller than women, due, in part, to sex-biased gene expression [34]), but typically pertain to anatomy, not cognition [35]. Prior work on sex differences in the auditory domain has focused largely on neural mechanisms in the periphery [36–39], although more recent work has demonstrated hormone-mediated differences in forebrain auditory responses to conspecific songs in birds [40,41]. Here, we show a small, but consistent sex difference in an auditory cognitive task – flexible rhythm pattern recognition – in a sexually dimorphic bird.

Yet just as some women are taller than some men, we find that some individual female zebra finches can outperform individual males in our tasks (e.g., probe tests in Fig. 3). How can this be reconciled with our hypothesis of a link between the neural circuitry for vocal learning and flexible rhythm pattern perception? Recent neuroanatomical work found that although female zebra finches possess smaller vocal motor regions, the overall network connectivity of vocal premotor regions is similar in male and female zebra finches [29]. Similarities in ascending auditory inputs to premotor regions, in particular, suggest that auditory processing mechanisms may be conserved across sex. This raises the question of whether individual variation in auditory-motor circuitry correlates with individual differences in flexibility of rhythmic pattern perception, irrespective of sex. Indeed, in humans, the strength of cortical auditory-motor connections predicts individual differences in rhythmic abilities [42,43].

Our finding that female zebra finches perform slightly worse than males in recognizing isochrony contrasts with a prior finding that female zebra finches outperform males in a task involving temporal processing of rhythmic signals [30]. In that study, birds (6 males and 6 females) called antiphonally with a robotic partner that emitted calls at a rate of 1 call/second, and quickly learned to adjust the timing of their calls in order to avoid a jamming signal introduced at a fixed latency after the robot call. While both studies highlight the importance of prediction for rhythm perception, the underlying mechanisms for predicting upcoming events may differ between the two studies. In principle, avoidance of the jamming signal could have resulted from learning a single temporal interval – the time between the robot call and the jamming signal. In contrast, in our task birds had to learn to recognize the relative timing of events and to respond to the same pattern, even when the absolute interval durations changed. Prior work suggests distinct mechanisms for single-interval versus and relative timing [12–14], so a male advantage on a task the depends on relative timing is not necessarily inconsistent with a female advantage in single interval timing.

Several recent studies have begun to elucidate the neural mechanisms for rhythm perception in nonhuman animals. For example, Asokan et al. (2021) found that the responses of neurons in the primary auditory cortex of mice are modulated by the rhythmic structure of a sound sequence. While neurons in the midbrain and thalamus encode local temporal intervals with a short latency, cortical neurons integrate inputs over longer a timescale (~1 s), and the timing of cortical spikes differs depending on whether consecutive sounds are arranged in a repeating rhythmic pattern or are randomly timed. In gerbils, midbrain neurons have also been shown to exhibit context-dependent responses: on average, neural responses were greater for noise bursts that occurred on the beat compared to when the same noise bursts occurred off-beat [45]. Similarly, in rhesus monkeys, occasional deviant sounds in auditory sequences elicit a larger auditory mismatch negativity in electroencephalogram recordings when those sequences had isochronous vs. arrhythmic event timing [46]. While these studies found context-dependent modulation of auditory responses, they did not test the ability of the animals to recognize a learned rhythm independently of tempo. Demonstrating this ability requires behavioral methods, and an important lesson from prior research is that training methods can strongly influence to what extent such abilities are revealed (e.g., compare [47] with [22] and [48] with [49]; see also [50]). More generally, the ability to relate neural activity to perception and behavior is critical for understanding the contributions of

motor regions to detecting temporal periodicity and predicting the timing of upcoming events, two hallmarks of human rhythm central to music's positive effect on a variety of neurological disorders.

Data accessibility. Concatenated trial data for female zebra finches, summarized data files, and female analysis code are available through Mendeley Data: https://dx.doi.org/10.17632/2r29x6gr7w.1

Published data for male zebra finches as well as stimulus files, circuit diagrams, male analysis code, and operant chamber setup diagram are available through Mendeley Data: https://dx.doi.org/10.17632/fw5f2vrf4k.2

The modified Pyoperant code for this experiment is available on GitHub: https://github.com/arouse01/pyoperant.

Acknowledgments. We thank T. Gardner and his laboratory for equipment and technical assistance with the operant chambers and T. Gentner for sharing the Pyoperant code. We thank members of the Kao lab for useful discussions and comments on earlier versions of this manuscript.

Funding. This work was supported by a Tufts University Collaborates grant (M.H.K. and A.D.P.), NIH Grant R21NS114682 (M.H.K. and A.D.P.), and a Canadian Institute for Advanced Research catalyst grant (A.D.P.).

References

1. Clemens J, Schöneich S, Kostarakos K, Hennig RM, Hedwig B. 2021 A small, computationally flexible network produces the phenotypic diversity of song recognition in crickets. *Elife* **10**, 1–31. (doi:10.7554/eLife.61475)

2. Mathevon N, Casey C, Reichmuth C, Charrier I. 2017 Northern Elephant Seals Memorize the Rhythm and Timbre of Their Rivals' Voices. *Current Biology* **27**, 2352-2356.e2. (doi:10.1016/j.cub.2017.06.035)

3. Garcia M, Theunissen F, Sèbe F, Clavel J, Ravignani A, Marin-Cudraz T, Fuchs J, Mathevon N. 2020 Evolution of communication signals and information during species radiation. *Nature Communications* **11**, 1–15. (doi:10.1038/s41467-020-18772-3)

4. Schöneich S, Kostarakos K, Hedwig B. 2015 An auditory feature detection circuit for sound pattern recognition. *Science Advances* **1**. (doi:10.1126/sciadv.1500325)

5. Schiavo JK, Valtcheva S, Bair-Marshall CJ, Song SC, Martin KA, Froemke RC. 2020 Innate and plastic mechanisms for maternal behaviour in auditory cortex. *Nature* **587**, 426–431. (doi:10.1038/s41586-020-2807-6)

6. Levitin DJ, Cook PR. 1996 Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception and Psychophysics* **58**, 927–935. (doi:10.3758/BF03205494)

7. Espinoza-Monroy M, de Lafuente V. 2021 Discrimination of Regular and Irregular Rhythms Explained by a Time Difference Accumulation Model. *Neuroscience* **459**, 16–26. (doi:10.1016/j.neuroscience.2021.01.035)

8. Savage PE, Brown S, Sakai E, Currie TE. 2015 Statistical universals reveal the structures and functions of human music. *Proc Natl Acad Sci U S A* **112**, 8987–8992. (doi:10.1073/pnas.1414495112)

9. Dalla Bella S, Benoit CE, Farrugia N, Keller PE, Obrig H, Mainka S, Kotz SA. 2017 Gait improvement via rhythmic stimulation in Parkinson's disease is linked to rhythmic skills. *Scientific Reports* **7**, 1–11. (doi:10.1038/srep42005)

10. Krotinger A, Loui P. 2021 Rhythm and groove as cognitive mechanisms of dance intervention in Parkinson's disease. *PLoS ONE* **16**, 1–20. (doi:10.1371/journal.pone.0249933)

11. Trehub SE, Thorpe LA. 1989 Infants' perception of rhythm: categorization of auditory sequences by temporal structure. *Canadian Journal of Psychology* **43**, 217–229. (doi:10.1037/h0084223)

12. Grube M, Cooper FE, Chinnery PF, Griffiths TD. 2010 Dissociation of duration-based and beat-based auditory timing in cerebellar degeneration. *Proceedings of the National Academy of Sciences* **107**, 11597–11601. (doi:10.1073/pnas.0910473107)

13. Teki S, Grube M, Kumar S, Griffiths TD. 2011 Distinct Neural Substrates of Duration-Based and Beat-Based Auditory Timing. *Journal of Neuroscience* **31**, 3805–3812. (doi:10.1523/JNEUROSCI.5561-10.2011)

14. Breska A, Ivry RB. 2018 Double dissociation of single-interval and rhythmic temporal prediction in cerebellar degeneration and Parkinson's disease. *Proc Natl Acad Sci U S A* **115**, 12283–12288. (doi:10.1073/pnas.1810596115)

15. Grahn JA, Brett M. 2007 Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience* **19**, 893–906. (doi:10.1162/jocn.2007.19.5.893)

16. Chen JL, Penhune VB, Zatorre RJ. 2008 Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex* **18**, 2844–2854. (doi:10.1093/cercor/bhn042)

17. Ross JM, Iversen JR, Balasubramaniam R. 2018 The Role of Posterior Parietal Cortex in Beat-based Timing Perception: A Continuous Theta Burst Stimulation Study. *Journal of Cognitive Neuroscience* **30**, 634–643. (doi:10.1162/jocn_a_01237)

18. Cannon JJ, Patel AD. In press. How beat perception coopts motor neurophysiology. *Trends in Cognitive Sciences* (doi:10.1101/805838)

19. Patel AD, Iversen JR. 2014 The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience* **8**, 57. (doi:10.3389/fnsys.2014.00057)

20. Roeske TC, Tchernichovski O, Poeppel D, Jacoby N. 2020 Categorical Rhythms Are Shared between Songbirds and Humans. *Current Biology* **30**, 3544-3555.e6. (doi:10.1016/j.cub.2020.06.072)

21. Norton P, Scharff C. 2016 “Bird song metronomics”: Isochronous organization of zebra finch song rhythm. *Frontiers in Neuroscience* **10**, 97–99. (doi:10.3389/fnins.2016.00309)

22. Rouse AA, Patel AD, Kao MH. 2021 Vocal learning and flexible rhythm pattern perception are linked: Evidence from songbirds. *Proceedings of the National Academy of Sciences* **118**, e2026130118. (doi:10.1073/pnas.2026130118)

23. Hulse SH, Humpal J, Cynx J. 1984 Discrimination and Generalization of Rhythmic and Arrhythmic Sound Patterns by European Starlings (Sturnus vulgaris). *Music Perception* **1**, 442–464. (doi:10.2307/40285272)

24. Rouse AA, Patel AD, Kao MH. 2021 Vocal learning and flexible rhythm pattern perception are linked: Evidence from songbirds. *Proceedings of the National Academy of Sciences* **118**, e2026130118. (doi:10.1073/pnas.2026130118)

25. Hagmann CE, Cook RG. 2010 Testing meter, rhythm, and tempo discriminations in pigeons. *Behavioural Processes* **85**, 99–110. (doi:10.1016/j.beproc.2010.06.015)

26. Celma-Miralles A, Toro JM. 2020 Discrimination of temporal regularity in rats (Rattus norvegicus) and humans (Homo sapiens). *Journal of Comparative Psychology* (doi:10.1037/com0000202)

27. Nottebohm F, Arnold AP. 1976 Sexual Dimorphism in Vocal Control Areas of the Songbird Brain. *Science (1979)* **194**, 211–213. (doi:10.1126/science.959852)

28. Kriengwatana B, Spierings MJ, ten Cate C. 2016 Auditory discrimination learning in zebra finches: effects of sex, early life conditions and stimulus characteristics. *Animal Behaviour* **116**, 99–112. (doi:10.1016/j.anbehav.2016.03.028)

29. Shaughnessy DW, Hyson RL, Bertram R, Wu W, Johnson F. 2019 Female zebra finches do not sing yet share neural pathways necessary for singing in males. *Journal of Comparative Neurology* **527**, 843–855. (doi:10.1002/cne.24569)

30. Benichov JI, Benezra SE, Vallentin D, Globerson E, Long MA, Tchernichovski O. 2016 The forebrain song system mediates predictive call timing in female and Male Zebra finches. *Current Biology* **26**, 309–318. (doi:10.1016/j.cub.2015.12.037)

31. Zann RA. 1996 *The zebra finch: a synthesis of field and laboratory studies*. Oxford, UK: Oxford University Press.

32. Gess A, Schneider DM, Vyas A, Woolley SMN. 2011 Automated auditory recognition training and testing. *Animal Behaviour* **82**, 285–293. (doi:10.1016/j.anbehav.2011.05.003)

33. van der Aa J, Honing H, ten Cate C. 2015 The perception of regularity in an isochronous stimulus in zebra finches (Taeniopygia guttata) and humans. *Behavioural Processes* **115**, 37–45. (doi:10.1016/j.beproc.2015.02.018)

34. Naqvi S, Godfrey AK, Hughes JF, Goodheart ML, Mitchell RN, Page DC. 2019 Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science (1979)* **365**. (doi:10.1126/science.aaw7317)

35. Zentner M. 2021 Social bonding and credible signaling hypotheses largely disregard the gap between animal vocalizations and human music. *Behavioral and Brain Sciences* **44**, e120. (doi:10.1017/S0140525X2000165X)

36. Krizman J, Bonacina S, Kraus N. 2020 Sex differences in subcortical auditory processing only partially explain higher prevalence of language disorders in males. *Hearing Research* **398**, 108075. (doi:10.1016/j.heares.2020.108075)

37. Krizman J, Rotondo EK, Nicol T, Kraus N, Bieszczad K. 2021 Sex differences in auditory processing vary across estrous cycle. *Scientific Reports* **11**, 1–7. (doi:10.1038/s41598-021-02272-5)

38. Berninger E. 2007 Characteristics of normal newborn transient-evoked otoacoustic emissions: Ear asymmetries and sex effects. *International Journal of Audiology* **46**, 661–669. (doi:10.1080/14992020701438797)

39. Gall MD, Salameh TS, Lucas JR. 2013 Songbird frequency selectivity and temporal resolution vary with sex and season. *Proceedings of the Royal Society B: Biological Sciences* **280**. (doi:10.1098/rspb.2012.2296)

40. Krentzel AA, Remage-Healey L. 2015 Sex differences and rapid estrogen signaling: A look at songbird audition. *Frontiers in Neuroendocrinology* **38**, 37–49. (doi:10.1016/j.yfrne.2015.01.001)

41. Brenowitz EA, Remage-Healey L. 2016 It takes a seasoned bird to be a good listener: Communication between the sexes. *Current Opinion in Neurobiology* **38**, 12–17. (doi:10.1016/j.conb.2016.01.005)

42. Blecher T, Tal I, Ben-Shachar M. 2016 White matter microstructural properties correlate with sensorimotor synchronization abilities. *Neuroimage* **138**, 1–12. (doi:10.1016/j.neuroimage.2016.05.022)

43. Vaquero L, Ramos-Escobar N, François C, Penhune V, Rodríguez-Fornells A. 2018 White-matter structural connectivity predicts short-term melody and rhythm learning in non-musicians. *Neuroimage* **181**, 252–262. (doi:10.1016/j.neuroimage.2018.06.054)

44. Asokan MM, Williamson RS, Hancock KE, Polley DB. 2021 Inverted central auditory hierarchies for encoding local intervals and global temporal patterns. *Current Biology* **31**, 1762-1770.e4. (doi:10.1016/j.cub.2021.01.076)

45. Rajendran VG, Harper NS, Garcia-Lazaro JA, Lesica NA, Schnupp JWH. 2017 Midbrain adaptation may set the stage for the perception of musical beat. *Proceedings of the Royal Society B: Biological Sciences* **284**. (doi:10.1098/rspb.2017.1455)

46. Honing H, Bouwer FL, Prado L, Merchant H. 2018 Rhesus monkeys (Macaca mulatta) sense isochrony in rhythm, but not the beat: Additional support for the gradual audiomotor evolution hypothesis. *Frontiers in Neuroscience* **12**, 475. (doi:10.3389/fnins.2018.00475)

47. van der Aa J, Honing H, ten Cate C. 2015 The perception of regularity in an isochronous stimulus in zebra finches (Taeniopygia guttata) and humans. *Behavioural Processes* **115**, 37–45. (doi:10.1016/j.beproc.2015.02.018)

48. Hulse SH, Humpal J, Cynx J. 1984 Discrimination and Generalization of Rhythmic and Arrhythmic Sound Patterns by European Starlings (Sturnus vulgaris). *Music Perception* **1**, 442–464. (doi:10.2307/40285272)

49. Samuels B, Grahn J, Henry MJ, Macdougall-Shackleton SA. 2021 European starlings (sturnus vulgaris) discriminate rhythms by rate, not temporal patterns. *J Acoust Soc Am* **149**, 2546. (doi:10.1121/10.0004215)

50. Bouwer FL, Nityananda V, Rouse AA, ten Cate C. 2021 Rhythmic abilities in humans and non-human animals: a review and recommendations from a methodological perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences* **376**. (doi:10.1098/rstb.2020.0335)

Figures

Figure 1. Experimental procedure for testing the ability to flexibly perceive rhythmic patterns. *(A)* Amplitude waveforms of isochronous (*Iso*) and arrhythmic (*Arr*) sequences of a repeated song element (sound B; see Fig. S1) with 120 ms (*top*) and 180 ms (*bottom*) mean inter-onset interval (IOI). *(B)* Schematic of the experimental procedure. After an initial pre-training procedure (not shown), each bird was trained to discriminate between isochronous and arrhythmic stimuli at 120 and 180 ms IOI, starting with sound A, followed by B and C ('ABC training', n=6 females) or starting with sound C followed by A and B ('CAB training', n=7 females) (color indicates sound type; see Fig. S1). To test for the ability to generalize the discrimination to new tempi, probe stimuli (144 ms IOI) were introduced after birds had successfully completed two training blocks. After training and probe testing with the three sound types, birds were trained with a broader stimulus set with a novel sound element (sound D) that included every integer rate between 75 and 275 ms IOI ("Rule Training"; see *Materials and Methods* for details). Color conventions indicating sound type are used in all subsequent figures. *(C)* Comparison of training time, plotted as Trials to Criteria, for male and female birds that completed rhythm discrimination training for all three blocks. Symbols denote sex. For this and subsequent figures, data from male birds were collected previously [22], and are replotted here for comparison. A Mann-Whitney test for each training block showed no sex-based differences in training time for any block ($p > 0.05$).

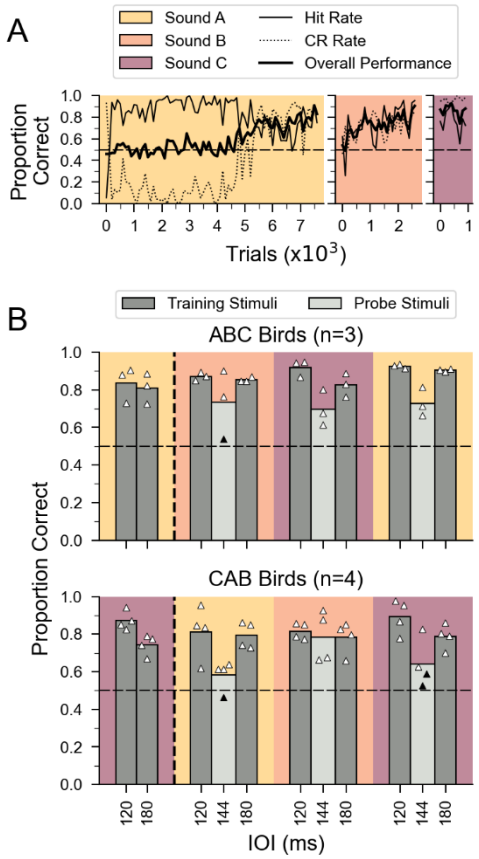

Figure 2. Learning curves for rhythmic pattern discrimination and generalization of discrimination to new tempi for female birds. (*A)* Performance of a representative female bird (*y7o97*) across three rhythm discrimination training blocks (ABC training) in 100-trial bins. *Thin black* line shows proportion correct for isochronous stimuli (S+, "hit rate"); *dotted* line shows proportion correct for arrhythmic stimuli (S-, "correct rejection"); *thick black* line indicates overall proportion of correct responses. Dashed horizontal line indicates chance performance. Data are plotted until performance reached criterion. (*B*) Training and probe test results for successful rhythm discrimination (n=7 female birds). Data to the left of the vertical dashed line indicate performance in the last 500 trials of the first rhythm discrimination training block (which did not include probe testing, see Fig. 1B). Data to the right of the vertical dashed line indicate performance during probe testing with stimuli at an untrained tempo of 144 ms IOI (*light gray*) and for interleaved training stimuli (*dark gray*). Triangles denote performance for each female (n = 3 probe tests/bird x 7 birds). All seven birds successfully generalized the isochronous versus arrhythmic pattern discrimination to novel tempi (20% from the two training tempi) for at least two sounds. Performance on probe stimuli was significantly above chance for 17 out of 21 probe tests (*white* triangles; $p<.0.0167$, binomial test with Bonferroni correction; black triangles denote performance not significantly different from chance). Bars represent average performance across birds in each group.

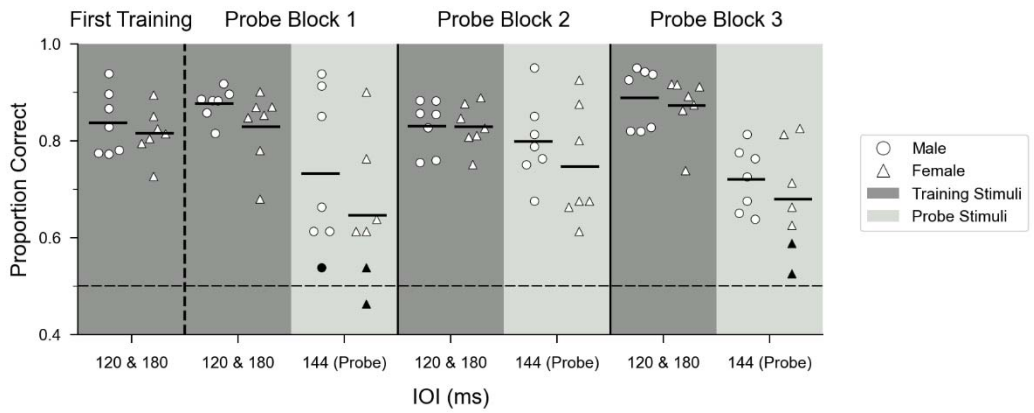

Figure 3. Comparison of generalization of rhythm discrimination by sex. Mean performance of all successful males (*circles;* n=7) and females (*triangles*, n=7) during the last 500 trials of the first training block (first column) and for probe (*light gray*) and interleaved training stimuli (*dark gray)* in subsequent blocks. Data are collapsed across training order, and interleaved training trials are combined across tempo. Dashed horizontal line indicates chance perform9nce. Filled black symbols are not significantly different from chance. Across probe tests, average performance was significantly lower for females than for males by ~9% (p<0.05, mixed-effects logistic regression).

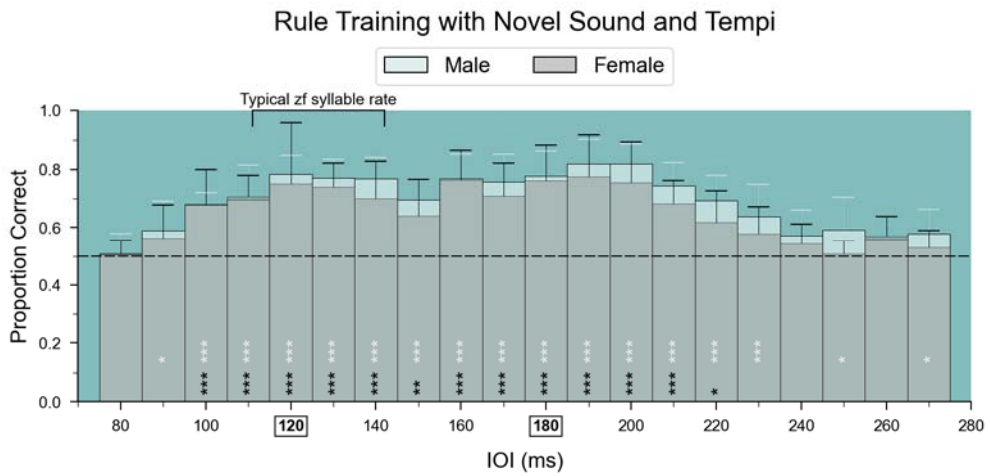

Figure 4. Sex differences in generalization of the learned discrimination across a wide range of tempi. Average performance of 5 female birds during the first 1000 trials of "Rule training" (dark gray bars). Bars represent performance in 10 ms bins; error bars denote SD. Dashed horizontal line indicates chance performance. Performance of females for binned IOIs between 95 ms and 225 ms are significantly different from performance for stimuli with 75-85 ms IOIs, which was not different from chance (***p<0.001; **p<0.005; *p<0.05, mixed-effect logistic regression, black asterisks). Data from 7 males from [22] are plotted for comparison (light green bars and white asterisks). Rates used in discrimination training prior to rule training are indicated in x-axis with bold and boxes. Color indicates sound type following conventions in Fig. 1B.

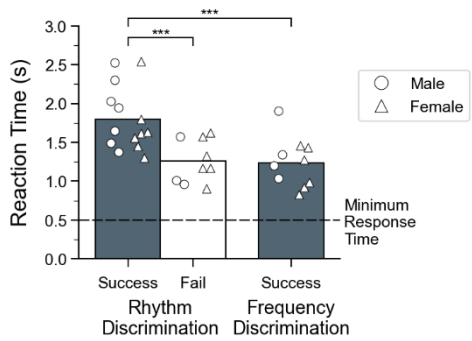

Figure 5. Reaction times during rhythmic pattern discrimination and frequency discrimination tasks. Mean response time for the last 500 responses in the first training block are shown for three groups of birds. For rhythm discrimination, data are shown for birds which completed all three training blocks (first bar, n=7 males and 7 females) and birds that did not complete rhythm discrimination training (middle bar, n=3 males and 6 females). For frequency discrimination, data are shown for a separate cohort of birds (third bar, n=4 males and 6 females). Horizontal line indicates the 0.5 s period between trial start and response switch activation when subjects could not make a response. Symbols indicate the average reaction time for each bird. A 2-way ANOVA of sex, group, and sex-by-group interaction showed only a significant effect of group ($F(2,27)$, $p = 0.001$). Successful rhythm discrimination birds responded significantly slower than the other two groups ($p<0.01$ compared to the failed birds, $p<0.005$ compared to the frequency discrimination birds; Tukey post-hoc tests).

Supplemental Figures

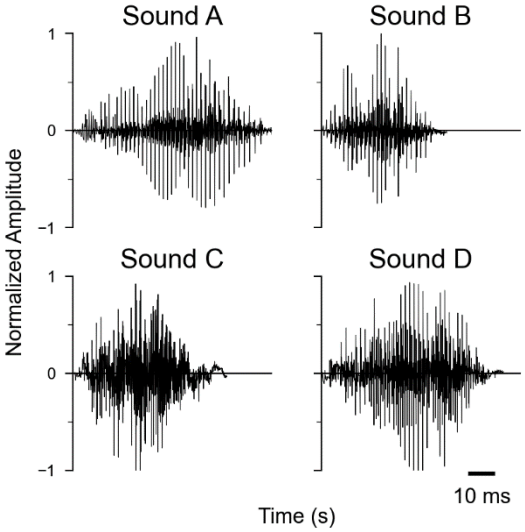

Figure S1. Sound element waveforms. Amplitude waveforms of the four sound elements used in the stimuli. Sound files are available at https://dx.doi.org/10.17632/fw5f2vrf4k.2

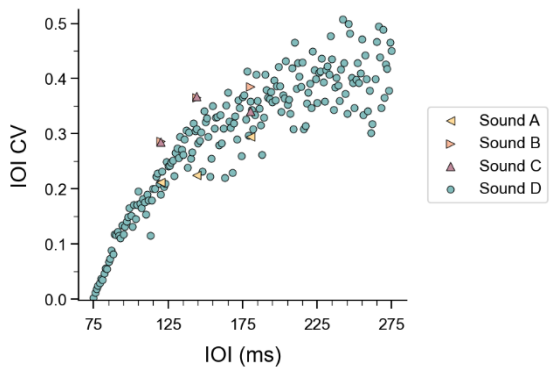

Figure S2. Variability in the timing of arrhythmic stimuli. Coefficient of variation (CV) of inter-onset-intervals in each arrhythmic stimulus plotted against stimulus mean IOI. As the average IOI of an arrhythmic stimulus approaches the length of the sound element used to make the sequence, variability in IOI duration approaches 0.

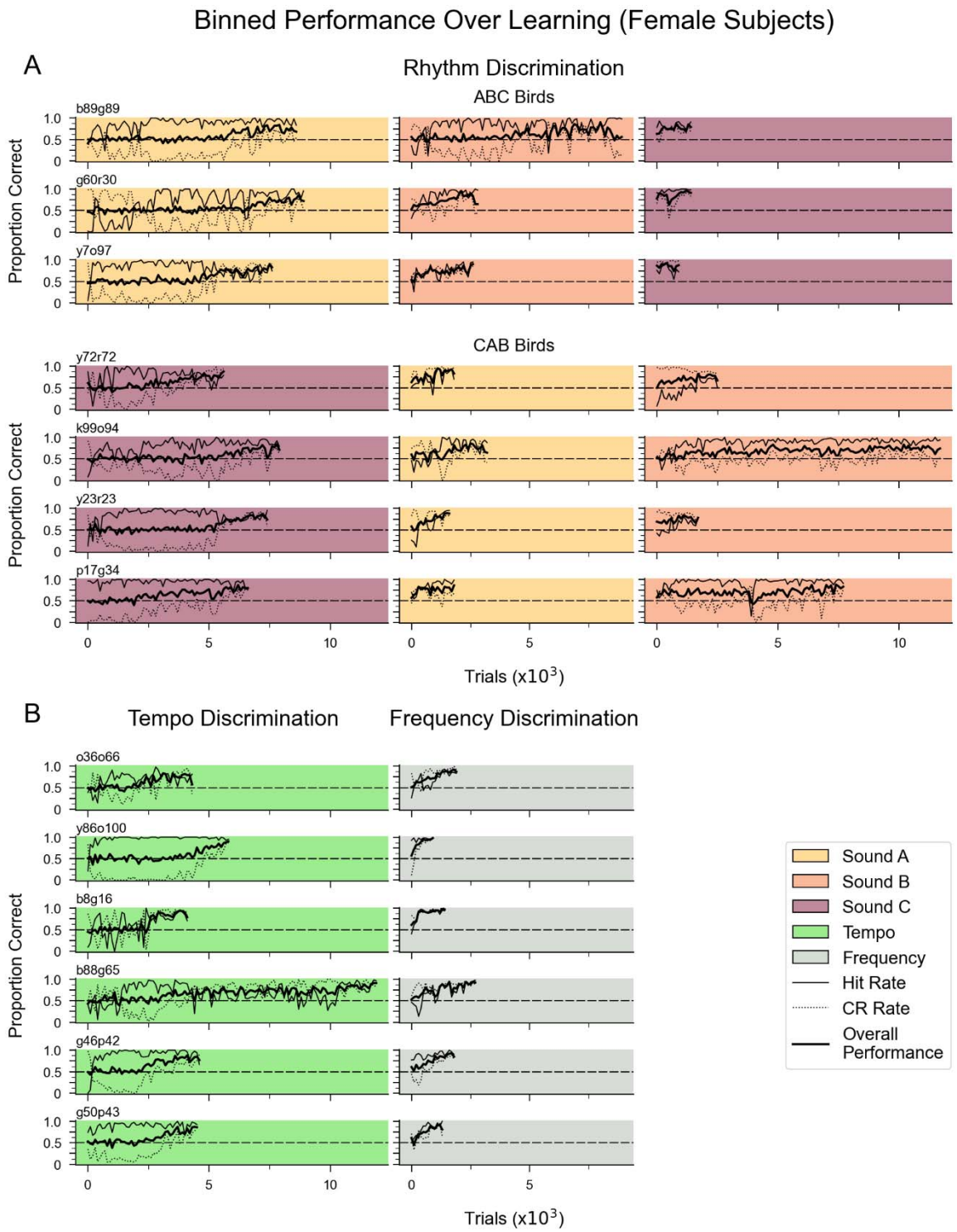

Figure S3. Individual learning curves for discrimination of stimuli based on rhythmic pattern, tempo, or frequency. (*A*) Learning curves for 7 female birds that successfully discriminated between isochronous and arrhythmic stimuli within 30 days per sound type. Thin black line indicates proportion correct for isochronous stimuli (S+, "hit rate"); dotted gray line indicates proportion correct for arrhythmic stimuli (S-, "correct rejection"); thick black line indicates overall proportion of correct responses. Dashed horizontal line indicates chance performance. Data are plotted until performance reached criterion (see *Materials and Methods*). Color corresponds to sound type (see Fig. 1B). (*B*) Learning curves for a

separate cohort of females (n=6) trained to discriminate between isochronous sequences of sound A that differed in tempo (120 ms IOI vs. 144 ms IOI) or frequency (shifted up 3 semitones vs. 6 semitones). All females tested met the criterion for discriminating stimuli based on tempo or frequency. Conventions as in (*A*).

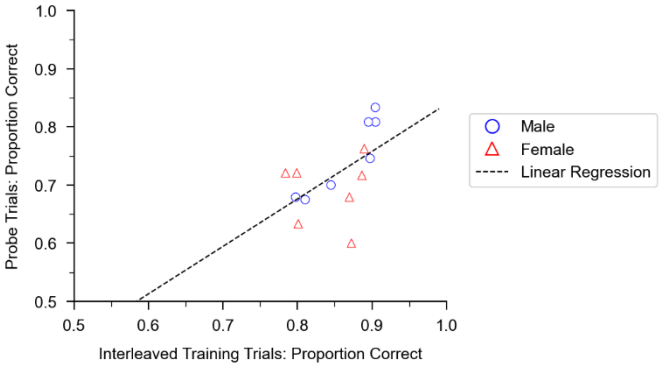

Figure S4. Relationship between performance on probe trials and interleaved training trials. Across birds that successfully completed all three training blocks (n=14), there was a significant positive correlation between the accuracy on interleaved training stimuli (learned discrimination) and probe stimuli (generalization); least squares regression fit; R^2 =0.31; p = 0.0375). For each bird, data were combined across all three sound elements; each data point is the average performance across 240 probe stimuli and ~2000 interleaved training stimuli.

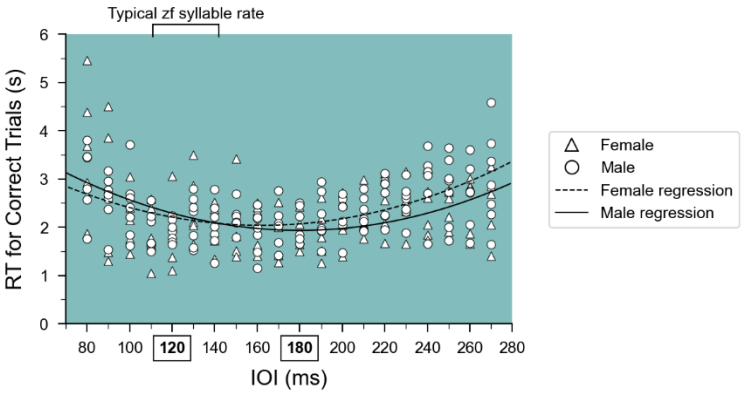

Figure S5. Reaction time (RT) across a broad range of tempi. Mean reaction time for correct trials versus tempo during Rule Training for male (*circles*) and female (*triangles*) birds. RT tended to be longer at the end of the tempo range (< 100 ms and > 220 ms), when discrimination of isochronous vs. arrhythmic stimuli fell to chance. Fitted quadratic curves are plotted to help illustrate the observed relationship.