



Vocal learning and flexible rhythm pattern perception are linked: Evidence from songbirds

Andrew A. Rouse^{a,1} , Aniruddh D. Patel^{a,b} , and Mimi H. Kao^{c,d}

^aDepartment of Psychology, Tufts University, Medford, MA 02155; ^bProgram in Brain, Mind and Consciousness, Canadian Institute for Advanced Research, Toronto, ON M5G 1M1 Canada; ^cDepartment of Biology, Tufts University, Medford, MA 02155; and ^dNeuroscience Graduate Program, Tufts University, Boston, MA 02111

Edited by Asif A Ghazanfar, Princeton University, Princeton, NJ, and accepted by Editorial Board Member Thomas D. Albright June 3, 2021 (received for review December 24, 2020)

Rhythm perception is fundamental to speech and music. Humans readily recognize a rhythmic pattern, such as that of a familiar song, independently of the tempo at which it occurs. This shows that our perception of auditory rhythms is flexible, relying on global relational patterns more than on the absolute durations of specific time intervals. Given that auditory rhythm perception in humans engages a complex auditory–motor cortical network even in the absence of movement and that the evolution of vocal learning is accompanied by strengthening of forebrain auditory–motor pathways, we hypothesize that vocal learning species share our perceptual facility for relational rhythm processing. We test this by asking whether the best-studied animal model for vocal learning, the zebra finch, can recognize a fundamental rhythmic pattern—equal timing between event onsets (isochrony)—based on temporal relations between intervals rather than on absolute durations. Prior work suggests that vocal nonlearners (pigeons and rats) are quite limited in this regard and are biased to attend to absolute durations when listening to rhythmic sequences. In contrast, using naturalistic sounds at multiple stimulus rates, we show that male zebra finches robustly recognize isochrony independent of absolute time intervals, even at rates distant from those used in training. Our findings highlight the importance of comparative studies of rhythmic processing and suggest that vocal learning species are promising animal models for key aspects of human rhythm perception. Such models are needed to understand the neural mechanisms behind the positive effect of rhythm on certain speech and movement disorders.

songbird | isochrony | comparative cognition | auditory perception | finch

The perception of rhythmic patterns in auditory sequences is important for many species, ranging from crickets and birds recognizing conspecific songs to humans perceiving phrase boundaries in speech or beat patterns in music. Animals vary widely in the complexity of the neural mechanisms underlying rhythm perception. Female field crickets, for example, have a circuit of five interneurons that form a feature detector for recognizing the stereotyped pulse pattern of male song (1). At the other extreme, human rhythm perception engages a complex circuit including cortical auditory and premotor regions, the basal ganglia, cerebellum, thalamus, and supplementary motor areas that enable recognition of a given rhythmic pattern across a wide range of tempi (2, 3). Growing evidence suggests that human rhythm perception relies on auditory–motor interactions even in the absence of movement (4–7). Given that vocal learning species have evolved neural adaptations for auditory–motor processing, including specializations of premotor and basal ganglia regions, and communicate using acoustic sequences that are often rhythmically patterned (2, 8–12), we hypothesize that such species are advantaged for flexible auditory rhythm pattern perception. This “vocal learning and rhythmic pattern perception hypothesis” focuses on perception in the absence of overt movement and is thus relevant to species that do not spontaneously synchronize movements to auditory rhythms.

A key test of flexibility in rhythm pattern perception is determining whether an individual can recognize a temporal pattern in

novel sequences with substantially different absolute time intervals than used during training. A simple example is recognition of isochrony, or equal time intervals between events, independent of absolute interval duration. Isochrony is a core feature of music, and humans easily recognize it as a global relational pattern between events, independent of tempi (13). Furthermore, isochrony has been reported as an underlying pattern in several forms of animal communication (8, 10, 14). Given the possible overlap in the circuitry for vocal learning and rhythm processing, our hypothesis predicts that vocal learners should be able to flexibly recognize isochrony in auditory sequences.

A variety of prior studies are consistent with our hypothesis that robust auditory–motor neural interactions in vocal learning species enable flexible rhythmic pattern perception. Vocal non-learning pigeons (*Columbia livia*) cannot learn to discriminate isochronous rhythms from arrhythmic sound patterns, though they can learn to categorize acoustic sequences based on the rate (versus rhythm) of events (15). Recently, rats (*Rattus norvegicus*), another vocal nonlearning species, were successfully trained to discriminate isochronous from arrhythmic sound sequences, but they showed only weak generalization when tested at novel tempi, thus exhibiting limited flexibility in auditory rhythm pattern perception (13). In contrast, in the same study, humans readily discriminated isochronous from arrhythmic sound sequences when tested at novel tempi (13). Prior work has also shown that starlings (*Sternus vulgaris*), a songbird that learns and increases its song repertoire throughout life, can learn to discriminate isochronous from arrhythmic sound

Significance

We can recognize the cadence of a friend’s voice or the rhythm of a familiar song across a wide range of tempi. This shows that our perception of temporal patterns relies strongly on the relative timing of events rather than on specific absolute durations. This tendency is foundational to speech and music perception, but to what extent is it shared by other species? We hypothesize that animals that learn their vocalizations are more likely to share this tendency. Here, we show that a vocal learning songbird robustly recognizes a basic rhythmic pattern independent of rate. Our findings pave the way for neurobiological studies to identify how the brain represents and perceives the temporal structure of auditory sequences.

Author contributions: A.A.R., A.D.P., and M.H.K. designed research; A.A.R. and M.H.K. performed research; A.A.R. analyzed data; and A.A.R., A.D.P., and M.H.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. A.A.G. is a guest editor invited by the Editorial Board.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: andrew.rouse@tufts.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2026130118/-DCSupplemental>.

Published July 16, 2021.

patterns and robustly generalizes this discrimination when tested at novel rates (16). Together, these studies suggest that vocal learners recognize isochrony as a global, relational pattern rather than by simply memorizing specific temporal interval patterns (17). It is unclear, however, how prevalent flexible isochrony perception is among vocal learners. To address this question, a particularly interesting species is the zebra finch (*Taeniopygia guttata*), the best-studied animal model of vocal learning. Here, we ask whether zebra finches exhibit flexible temporal pattern perception. Male zebra finches (but not females) learn to sing temporally precise, hierarchically organized songs with structure on multiple timescales (18). Thus, we focus on the male zebra finch's ability to recognize isochrony as a global temporal feature of sound sequences. This rhythmic pattern is a common feature in zebra finch song (10, 11), making them a good choice for rhythm research.

Several lines of evidence suggest that zebra finches may be adept at flexible rhythm perception. First, like humans, zebra finches possess recurrent connections between auditory and motor regions specialized for learned vocalizations, and ablation of the motor-to-auditory pathway prevents the adaptive modification of song timing (19). Second, neurons in auditory association areas exhibit differential activation in response to hearing isochronous versus arrhythmic stimuli (20). More recently, predictive activity in the song motor pathway prior to anticipated calls has been reported during social call exchanges (21). Finally, several studies have shown that zebra finches can predict the timing of isochronous, antiphonal calls during vocal turn taking and can adjust their own call timing to avoid interference, an ability that is disrupted by manipulations of the song motor pathway (21–23). Taken together, these results are consistent with the notion that interactions between vocal motor and auditory regions may generate predictive timing signals that facilitate flexible rhythm perception.

In contrast to the hypothesis that vocal learners are adept at flexible rhythm perception, other studies suggest limited rhythm perception abilities in zebra finches. Using an operant conditioning paradigm, ten Cate and colleagues found that zebra finches can learn to discriminate isochronous from arrhythmic patterns but show weak generalization when tested at novel tempi (9, 24). This led them to propose that zebra finches attend to specific, local features of temporal patterns, such as the exact duration of individual temporal intervals, rather than to global temporal structure. However, several factors may have contributed to underestimating the rhythm perception abilities of zebra finches. These include the use of artificial sounds and training with just one pair of isochronous and arrhythmic stimuli (or with short sequences), which may have resulted in attention to local features of the stimuli.

Here, we take a different approach, employing longer temporal patterns composed of conspecific sounds and training birds sequentially using multiple sound types at different tempi. We first confirm that male zebra finches can discriminate isochronous from arrhythmic rhythms. Strikingly, in contrast to previous work, we find that male zebra finches can robustly generalize this discrimination to novel tempi distant from the training tempi, consistent with the idea that they can recognize rhythmic regularity based on global temporal patterns. This makes zebra finches a promising candidate for human-relevant neural research on rhythm perception, including the role of the motor system in rhythmic pattern perception. In particular, evolutionary parallels between avian and human vocal learning circuitry (12) make songbirds a tractable model for investigating the role of premotor and basal ganglia circuits in rhythm perception (25) and for exploring how and why rhythm processing deficits are associated with a variety of speech and motor disorders (26–29).

Results

To encourage the formation of a conceptual category of temporal regularity, we trained young male zebra finches in the late stage of sensorimotor learning (age: 62 to 106 d posthatch [dph] on the

first day of training) on rhythm discrimination using a modified go/no-go task (go/interrupt) with multiple training phases (Fig. 1*A* and *D*). To increase the likelihood that the stimuli were salient for the birds, we used sequences of natural stimuli: either an introductory element that is typically repeated at the start of a song or a short harmonic stack commonly sung by zebra finches (18) (Fig. 1*B*). In each training phase, birds were presented with an isochronous and an arrhythmic sequence of a single sound element, at two different tempi (120 and 180 ms interonset interval [IOI]; e.g., Fig. 1*C*; *Materials and Methods*). The average duration of the sequences was 2.32 ± 0.1 (SD) s (20 intervals and 13 intervals for the two training tempi, respectively), and the temporal pattern for each arrhythmic stimulus was unique. Once a bird successfully discriminated the isochronous versus arrhythmic stimuli in one phase (*Materials and Methods*), he underwent further training in a subsequent phase with a new set of isochronous and arrhythmic sequences with a novel song element (Fig. 1*D*). To encourage attention to temporal rather than spectral features of the stimuli, probe testing began after two successful training phases. To ensure that the particular training sequence did not matter, two different orderings of stimuli were used for rhythm discrimination training ($n = 4$ birds trained with sound A, then B, then C [“ABC birds”] and $n = 6$ trained using the order C, A, B [“CAB birds”]).

Using this training procedure, seven out of 10 male birds learned to discriminate isochronous versus arrhythmic sequences of each of three different song elements within 30 d per element. Fig. 2*A* shows the time course of learning for one representative ABC bird. Initially, this male responded to both the isochronous and arrhythmic stimuli but then gradually learned to withhold his response to the arrhythmic stimulus. Across all seven birds, the rhythm discrimination performance was ~84% accurate at the end of the first training block (Fig. 2*B*, left column), and birds maintained a comparable accuracy level by the end of training with all three stimuli (see *SI Appendix*, Fig. S1*A* for learning curves for each bird).

After successful completion of two phases of rhythm discrimination with two different sound elements (Fig. 1*D*), we tested whether male zebra finches could generalize the discrimination of isochronous versus arrhythmic stimuli to a novel tempo. Probe stimuli consisted of an isochronous sequence and an arrhythmic counterpart at a new tempo: 144 ms IOI, 20% faster than the slower training stimuli and 20% slower than the faster training stimuli. This tempo was chosen because we found that zebra finches can tell apart isochronous stimuli when the tempo differs by 20% (*SI Appendix*, Fig. S1*B*; $n = 4$ birds, mean performance = 84% correct at the completion of training; $P < 0.0001$, binomial test). Probe stimuli were presented in 10% of the trials, randomly interleaved with training stimuli, and responses to probes were never reinforced (*Materials and Methods*). Fig. 2*B* shows the performance on training stimuli and probe stimuli for the seven birds that successfully completed all three phases of rhythm discrimination training. For each song element (A, B, or C), the overall proportion of correct responses is plotted for 80 nonreinforced probe trials (light gray bars; equal probability of isochronous and arrhythmic sequences). Data for interleaved, training stimuli are shown for comparison (dark gray bars). Performance on probe stimuli was significantly above chance for 20 of 21 probe tests ($n = 3$ probe tests/bird \times 7 birds, $P < 0.0167$, binomial test with Bonferroni correction), indicating that the birds readily distinguished between the isochronous and arrhythmic stimuli at the new tempo (see also *SI Appendix*, Fig. S2 for graphs in Fig. 2*B* plotted as hit–false alarm rate).

The reaction times of birds to training stimuli indicate that they heard multiple intervals before responding. On average, the seven birds that successfully completed all three phases of rhythm discrimination training responded ~1.8 s after stimulus onset, or ~80% of the duration of the stimulus train (Fig. 3). In contrast, the average reaction time for the three birds that did not reach the

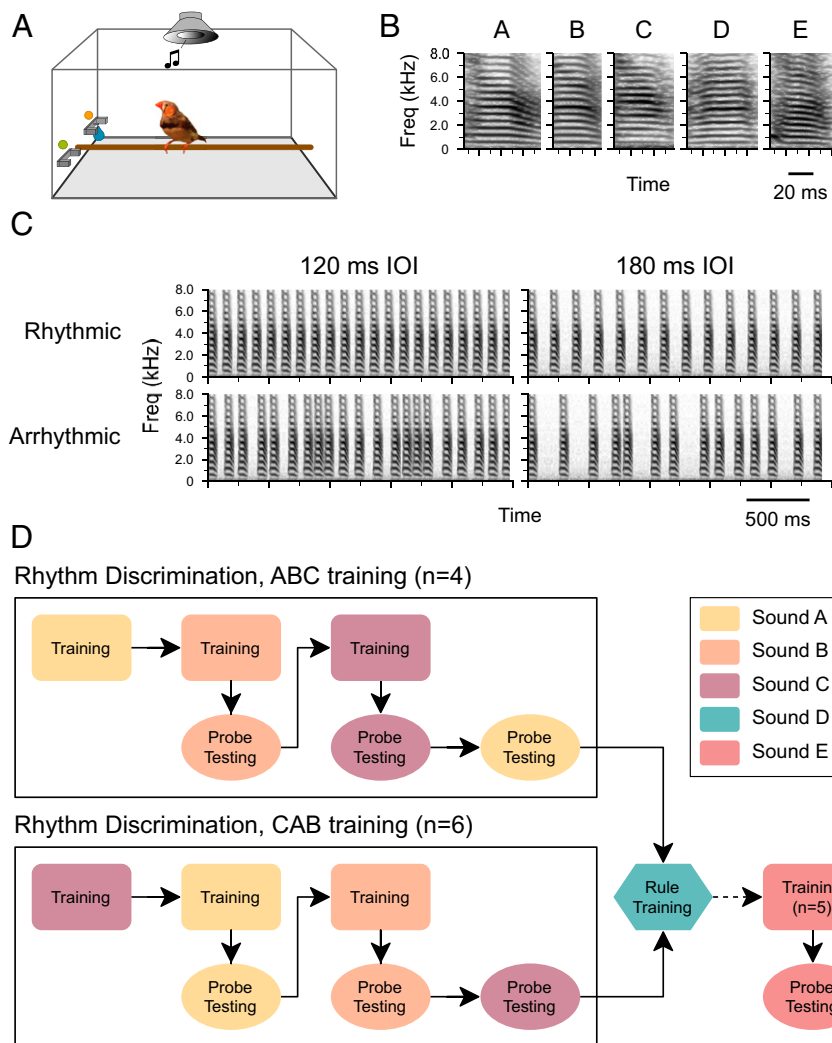


Fig. 1. Experimental procedure. (A) Birds were housed individually in small cages with trial and response switches and a speaker mounted overhead for sound playback. A water reward dispenser (blue droplet) was embedded in the response switch. (B) Song elements used to make rhythmic sequences included short stacks (A, C, and D) or introductory elements (B, E) (mean duration ~ 70 ms). (C) Spectrograms of isochronous (Top) and arrhythmic (Bottom) sequences with 120 or 180 ms mean IOI using sound A (Sound Files S1–S4). (D) Schematic of the experimental procedure. After an initial pretraining procedure (not shown), each bird was trained to discriminate between isochronous and arrhythmic stimuli, starting with sound A and followed by B and C (“ABC training”; $n = 4$) or starting with sound C followed by A and B (“CAB training”; $n = 6$) (color indicates sound type). Probe stimuli (144 ms IOI) were introduced after birds had successfully completed two phases of training. After training and probe testing with the three sound types, birds were trained with a broader stimulus set with a novel sound (sound D) that included every integer rate between 75 and 275 ms IOI (“rule training”; Materials and Methods) followed by a final training and testing phase with a novel sound (sound E). The colors indicating sound type are preserved in subsequent figures.

criterion for rhythm discrimination during training was ~ 1.2 s, $\sim 50\%$ of the duration of each stimulus. Thus, success versus failure at rhythm discrimination training may be related to how long birds listened to the stimuli before making a response, although more data would be needed to test this assertion statistically.

Rule Training and Generalization. The seven birds that completed rhythm discrimination training were then advanced to “rule training” in which they were presented with isochronous and arrhythmic stimuli over a wide range of 201 different tempi (75 to 275 ms IOI in 1 ms steps; Fig. 1D) using a new song element (sound D, duration = 74.7 ms; Fig. 1B). All responses were reinforced, but the tempi now extended well beyond the trained range. On a given trial, any one of 402 different stimuli was presented to a bird (equal probability of isochronous and arrhythmic stimuli at any tempo, with each arrhythmic stimulus being a unique temporal pattern). This made it less likely that birds could do rhythm discrimination

based on memorizing specific stimuli. We reasoned that if a subject had learned the “rule” of responding differentially to isochronous and arrhythmic patterns rather than discriminating patterns by memorizing specific stimuli, then the discrimination performance should be maintained across a wide range of rates. Fig. 4A shows the performance of seven males during the first 1,000 trials of rule training (~ 50 trials per bird per 10 ms IOI bin). As expected, performance fell to chance at very fast IOIs (75 to 85 ms IOI) in which the degree of temporal variation of intersyllable intervals was severely limited by the duration of the sound element (SI Appendix, Fig. S3 and Materials and Methods). Notably, performance on reinforced stimuli was significantly better than performance at the fastest bin for IOI bins between 95 and 235 ms (Fig. 4A; $P < 0.001$, mixed-effect logistic regression), which include tempi $\sim 30\%$ slower and $\sim 25\%$ faster than the original training range and beyond the typical range of zebra finch song tempo (18). These results suggest that performance was based on a principle of regularity rather than

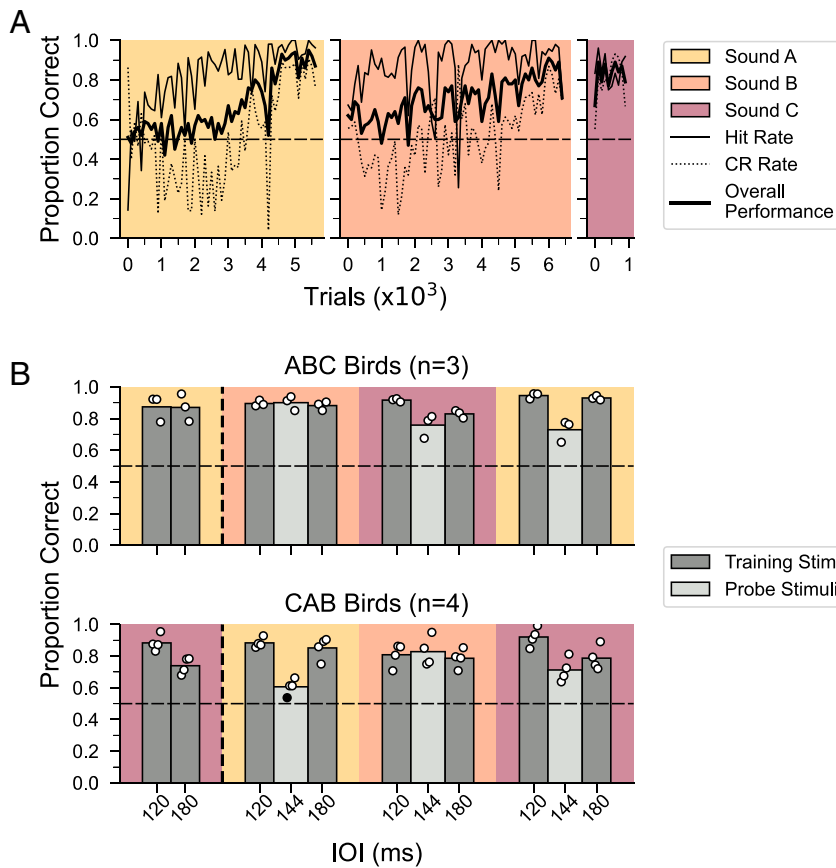


Fig. 2. Learning curves for rhythmic pattern discrimination and generalization of discrimination to new tempi. (A) Performance of a representative bird (b61g61) across three rhythmic discrimination training phases (ABC training) in 100-trial bins. The thin black line shows the proportion correct for isochronous stimuli (S+, “hit rate”), the dotted line shows the proportion correct for arrhythmic stimuli (S-, “correct rejection”), and the thick black line indicates the overall proportion of correct responses. The dashed horizontal line indicates chance performance. Data are plotted until performance reached criterion. (B) Training and probe test results for successful regularity discrimination ($n = 7$ out of 10 birds). Data to the left of the vertical dashed line indicate performance in the last 500 trials of the first rhythmic discrimination training phase (which did not include probe testing, cf Fig. 1D). Data to the right of the vertical dashed line indicate performance during probe testing for interleaved training (dark gray) and probe (light gray) stimuli. All seven males successfully generalized the isochronous versus arrhythmic tempo discrimination to novel tempi (20% from the two training tempi). The filled circles denote performance for each bird, the white circles denote performance significantly different from chance ($P < 0.00167$, binomial test with Bonferroni correction), and the black circle is not significantly different from chance. Bars represent average performance across birds in each group.

on rote memorization of stimuli. Performance remained stable with additional rule training trials (SI Appendix, Fig. S4).

To test whether the additional training strengthened the formation of a conceptual category of temporal regularity, we next asked whether a subset of birds ($n = 5$ of 7 birds) could generalize rhythm discrimination to sequences of a novel song element after rule training. Fig. 4B shows the individual learning curves for five birds trained with sequences of a novel song element (sound E; Fig. 1B). For all five birds, overall correct performance exceeded 75% within 1,000 trials, and each bird reached the criterion for successful discrimination in the minimum possible time (2 d). Moreover, performance on probe trials was significantly above chance (Fig. 4B, Right; $P < 0.0001$, binomial test; 82.5% correct for 80 probe stimuli versus 89% correct for interleaved training stimuli). Thus, after multiple phases of rhythm discrimination training with different sound elements, male birds quickly discriminated and categorized isochronous and arrhythmic sequences of novel stimuli.

Discussion

Humans readily discriminate between different temporal patterns and recognize the same patterns at different tempi, an ability that is present early in development (13, 30). Motivated by the finding that the perception of auditory rhythms engages forebrain

auditory–motor loops in humans, we hypothesized that the specialized auditory–motor forebrain circuitry that subserves vocal learning confers advantages in flexible rhythm pattern perception (4, 31–33). Using a sequential training paradigm with multiple sound types, timbres, and tempi, we show that a vocal learning songbird can readily learn to discriminate auditory patterns based on rhythm. Consistent with prior work, we found that male zebra finches can differentiate isochronous from arrhythmic patterns (9, 24). Strikingly, and unlike previous studies, this ability generalized to novel tempi distant from the training tempi, showing flexible perception of rhythmic patterns. When tested with stimuli at a tempo 20% faster or slower than the training stimuli, zebra finches robustly discriminated isochronous from arrhythmic patterns (Fig. 2B and SI Appendix, Fig. S2). Furthermore, correct discrimination of isochronous versus arrhythmic stimuli remained significantly above chance at tempi ranging from 30% faster to 25% slower than the original training stimuli (Fig. 4A and SI Appendix, Fig. S4). Together, these results demonstrate that zebra finches resemble humans in recognizing a fundamental auditory rhythmic pattern—isochny—based on global temporal patterns, independent of the absolute durations of specific intervals.

While our hypothesis predicts that vocal learning species are generally advantaged in flexible rhythm pattern perception, zebra

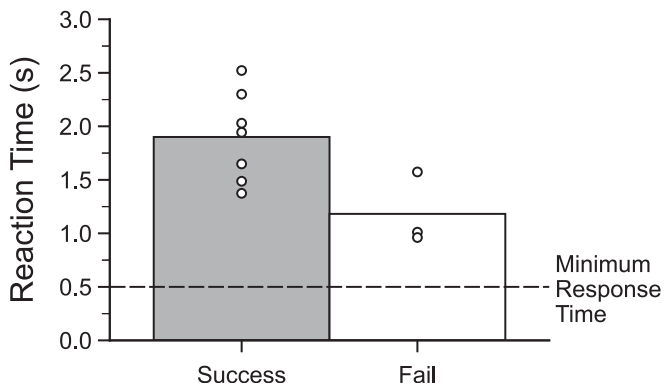


Fig. 3. Reaction times of successful and unsuccessful discrimination of isochronous versus arrhythmic stimuli. Mean response time for the last 500 responses in each training phase for birds trained to discriminate sequences based on rhythmic pattern (left bar: $n = 7$ birds which completed training and probe testing; right bar: $n = 3$ birds which did not reach performance criterion during training). The horizontal line indicates the 0.5 s time between the trial start and response switch activation when subjects could not make a response. Data points indicate the average reaction time for each bird; the conventions are as in Fig. 2.

finches may be specifically predisposed to attend to isochrony as a rhythmic pattern, given that a tendency toward isochrony is a feature of zebra finch song (10, 11). This is consistent with the idea that the structure of biologically relevant vocalizations shapes auditory

processing (34–36). Indeed, our finding that zebra finches can readily detect isochrony across a range of tempi may help explain how birds can recognize songs as coming from the same individual, even when the overall song duration is compressed or stretched by >25% (37).

Differences in experimental paradigms may account for the disparity between our findings and prior work that suggested that male zebra finches attend to local temporal features, such as the duration of single intervals, rather than to global temporal structure. First, prior studies used artificially constructed sequences of pure tones (9) or simple percussive sounds [woodblock (24)]. In contrast, we used species-specific sounds that may be more salient to the birds. Second, to encourage attention to temporal rather than spectral features of the stimuli, we used a novel sequential training procedure in which the tempi were kept constant, but different sound elements were presented in each phase (Fig. 1D). Across training phases, the stimuli differed in their spectral properties and intensity (SI Appendix, Fig. S3B), but birds were rewarded for successfully discriminating isochronous from arrhythmic stimuli regardless of frequency or intensity. Finally, rather than testing adult birds whose songs had stabilized, our study used younger male zebra finches that were actively using auditory feedback to modify their vocalizations in order to produce a good match to the memorized song model (62 to 106 dph on the first day of shaping). Taking into account such factors will be important for future studies probing the rhythm perception capacities of nonhuman animals.

Our finding that a vocal learning songbird exhibits flexible rhythm pattern perception contrasts sharply with findings in vocal nonlearning rats. A recent study found that rats can be trained to

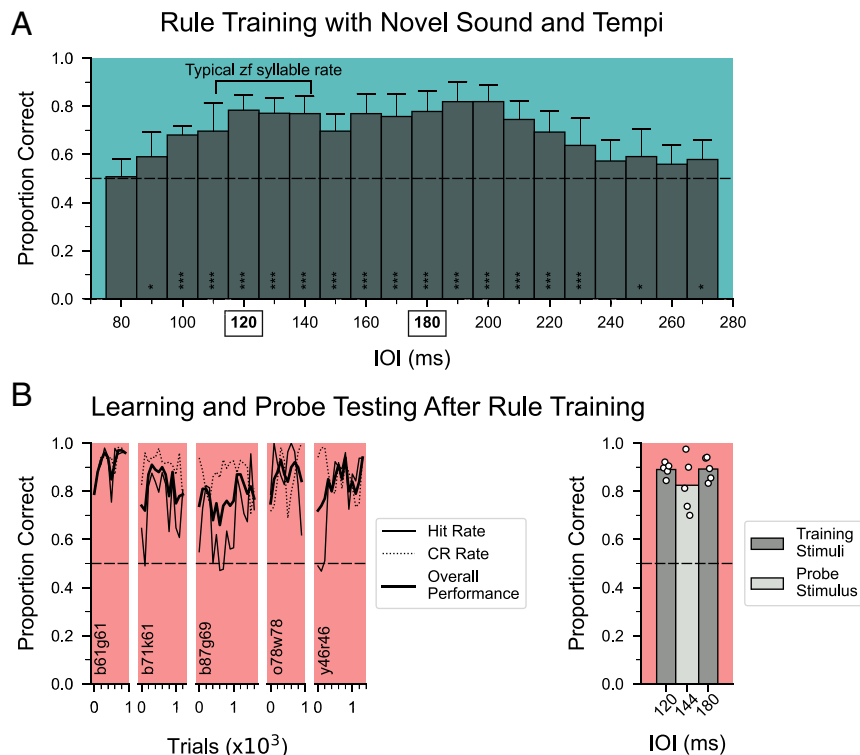


Fig. 4. Rule training results and subsequent training and testing results with a novel stimulus. (A) Average performance of the first 1,000 trials of rule training across birds that completed rhythmic pattern discrimination training ($n = 7$). Bars represent performance in 10-ms bins; error bars denote SD. The dashed horizontal line shows chance performance. Performance for IOIs between 95 and 235 ms are significantly different from performance during the 75 to 85 ms IOIs, which was not different from chance ($***P < 0.001$; $*P < 0.05$, mixed-effect logistic regression). The rates used in discrimination training prior to rule training are indicated in the x-axis with bold and boxes. (B, Left) Learning curves of five birds that learned a novel stimulus set (sound E) after completing rule training, plotted in 100-trial bins. All five learned the stimuli to criteria in the minimum possible time. (Right) Performance during probe testing for sound E was significantly different from chance for all five birds ($P < 0.001$, binomial test). The conventions are as in Fig. 2. In both A and B, sound type is indicated by the color following the conventions in Fig. 1D.

respond differently to isochronous versus arrhythmic stimuli but show weak generalization when tested at novel tempi (maximally ~15% different from training tempi). Specifically, rats responded only 5% more often to isochronous patterns at novel tempi (13). In contrast, the zebra finches in our study typically responded twice as much or more (median = 266% more often) to isochronous than to arrhythmic patterns at a novel tempo 20% different from training tempi, akin to human performance (13). This suggests that rats may be biased to attend to absolute stimulus durations in their perception of rhythms, while zebra finches more readily perceive global temporal patterns based on the relative timing of events. It remains to be seen if zebra finches outperform rats on tests of rhythm perception if the two species are trained and tested in comparable ways, which we predict based on differences in their vocal learning ability. More generally, the ability to uncover a species' rhythmic processing capacities may depend on the specific training and testing methods, which can substantially influence how an animal performs on tests of rhythmic processing (16, 38, 39).

Pertinent to our hypothesis, recent theoretical and empirical work suggests that vocal learning should not be considered a binary behavioral trait (40, 41). Rather, there may be a continuum in vocal learning capacities across species, comprised of multiple behavioral modules that may be targeted independently by evolutionary pressures. These include 1) the ability to coordinate the timing of vocalizations (e.g., antiphonal calling and duetting), 2) the ability to dynamically vary the acoustic properties of vocalizations (vocal production variability), and 3) the ability to adaptively modify vocalizations as a function of social and auditory experience (vocal plasticity). Across the spectrum of vocal learning abilities, there are differences in the extent of cross talk between forebrain auditory regions and forebrain motor regions that send signals to brainstem vocal pattern-generating circuits (42, 43). We hypothesize that the degree of sensorimotor connectivity correlates with differences in the ability to flexibly perceive auditory rhythms. Species that rely on auditory input to develop and maintain their vocalizations should have more flexible rhythmic pattern perception (e.g., of isochrony), reflecting the strength and plasticity of auditory-motor forebrain circuitry.

Specifically, our hypothesis predicts that, if trained and tested in comparable ways, species with greater vocal flexibility will exhibit faster learning rates for rhythmic discrimination and/or a greater degree of generalization to novel stimuli with different absolute time intervals than those used in discrimination training. For example, we predict that "singing mice" (*Scotinomys teguina*), which coordinate antiphonal songs to minimize temporal overlap, would be better at recognizing a rhythmic pattern independent of tempo compared to laboratory mice (*Mus musculus*), which do not exhibit vocal turn taking (44, 45), and marmosets (*Callithrix jacchus*) and Japanese macaques (*Macaca fuscata*), which call antiphonally with conspecifics (46, 47), will have more flexible rhythm pattern perception than rhesus macaques (*Macaca mulatta*), which do not. Similarly, we would predict that bat species that alter the timing of their vocalizations and produce isochronous calls or clicks for echolocation (8, 48, 49) would be better at discriminating isochronous versus arrhythmic stimuli than nonecholocating species. Comparative work could also be performed with related pinniped species that differ in the degree of vocal plasticity (50, 51). A particularly interesting comparison would be testing the rhythm perception abilities of vocal learning harbor seals (*Phoca vitulina*) (52) or gray seals (*Halichoerus grypus*) (53) versus the much less vocally flexible California sea lion (*Zalophus californianus*) (51). While there is good evidence that a California sea lion can flexibly entrain her movements to an isochronous auditory beat (54, 55), we predict that the seals would outperform sea lions in purely perceptual tests of rhythm.

Even among vocal learning species, our hypothesis predicts differences in the flexibility of rhythm perception capacities depending

on the degree of plasticity in the vocal learning circuitry. Thus, for example, birds that can learn new vocalizations throughout life ("open-ended learners") should outperform those whose song learning is limited to an early sensitive period ("closed-ended learners") (56). Similarly, it would be interesting to characterize rhythm perception abilities in birds whose vocal learning abilities vary seasonally (e.g., canaries, *Serinus canaria*) (57). Finally, our hypothesis also makes predictions regarding sexual dimorphism in rhythm perception abilities. In zebra finches and many other songbirds, only the males sing and possess pronounced forebrain motor regions for vocal production. Thus, we predict that in such species, males would outperform females in our rhythm discrimination and generalization tasks. Interestingly, a previous study showed that both male and female zebra finches can predict the timing of a partner's calls during antiphonal calling (22). It remains to be seen, however, how males and females compare in terms of their global rhythm pattern perception abilities.

One key issue for future research is elucidating the mechanisms by which isochronous patterns are recognized independent of tempo. Prior work has shown that oscillatory activity in the auditory cortex is differentially modulated when acoustic patterns switch from random sequences to predictably timed sequences, suggesting that cortical oscillations may play a role in predictive timing (58). More recently, single unit recordings in mice found that while subcortical neurons encode local temporal intervals with high fidelity, neurons in the primary auditory cortex are sensitive to stimulus regularity, a global rhythmic pattern (59). Distinct neural responses to a specific rhythmic pattern [as demonstrated in rats for isochrony (58)], however, are no guarantee that the pattern is robustly recognized at different tempi (13), which can only be shown with behavioral studies of rhythm perception and categorization.

A limitation of the current work is that we do not know if the finches' performance is based on retrospective or predictive mechanisms. Detection of isochrony could rely on a strictly retrospective mechanism in which the current interval is compared with the previous interval (17). Alternatively, isochrony detection may depend on predictions of the timing of future events based on the durations of prior intervals (60, 61). Current evidence in primates suggests that predictions of the timing of future events may arise in premotor areas, even in the absence of movements aligned to rhythmic sounds (5, 31). In songbirds, a premotor nucleus that controls song timing (25) has reciprocal connections with forebrain auditory regions (19, 62). Predictive activity in this region has been reported prior to anticipated calls with a vocal partner (21), and damage to or inactivation of the motor pathway interferes with temporally precise vocal turn taking (22, 23). These results are consistent with the notion that interactions between vocal motor and auditory regions may generate predictive timing signals that facilitate flexible rhythm perception. Future work disrupting motor region activity during the auditory processing of rhythms will help elucidate what role they play in rhythm perception and in shaping auditory cortical responses to rhythmic patterns.

Moving beyond isochronous patterns, a critical issue for future research is determining whether songbirds, like humans, are capable of abstracting an underlying periodicity from nonisochronous rhythms ("beat perception") (5). In our study, male zebra finches showed evidence of relative timing perception (i.e., for perceiving relations between durations that make up a rhythm), allowing them to generalize categorization of isochronous versus arrhythmic stimuli across salient changes in tempo. Relative timing perception is a prerequisite for beat perception but does not guarantee it. Determining whether zebra finches can perceive beats will require testing with more complex temporal patterns (63).

While birds and humans are distantly related, there are numerous parallels in their vocal learning capacities and underlying neural circuitry (12). There is growing interest in the brain mechanisms of

rhythm perception. This is driven in part by evidence that deficits in rhythm processing are linked to a number of childhood language disorders, including dyslexia, developmental language disorder, and stuttering (29), and by the observation that musical rhythm can facilitate movement in patients with basal ganglia movement disorders such as Parkinson's disease (28). The neural mechanisms underlying these findings are not well understood. Their elucidation would benefit from an animal model with flexible rhythm perception which possesses specialized auditory–motor forebrain circuitry allowing fine-grained, circuit-level measurements and manipulations.

Materials and Methods

Subjects. Subjects were 16 experimentally naïve male zebra finches from the Tufts University breeding colony (mean age = 73 ± 11 [SD] dph at the start of training; range = 62 to 106 dph). All procedures were approved by the Tufts University Institutional Animal Care and Use Committee.

Auditory Stimuli. All stimuli were assembled in MATLAB. Two elements common to zebra finch song—an introductory note (Fig. 1B, sounds B and E) and a short harmonic stack (Fig. 1B, sounds A, C, and D)—were selected from recordings of four unfamiliar birds (duration: 51 to 80 ms). For each song element, isochronous stimuli were generated by repeating the sound at a target rate until the stimulus was approximately ~ 2.3 s long (e.g., Fig. 1C). For every isochronous stimulus, a unique arrhythmic stimulus was generated with the same average IOI between syllables (e.g., Fig. 1C, maximum difference between the average IOIs for an isochronous stimulus and an arrhythmic stimulus at the same tempo was < 0.02 ms). 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.

Arrhythmic stimuli were generated in an iterative process using MATLAB. Briefly, as the number of intervals was known for the isochronous stimulus, an equal number of random intervals was generated, and then the arrhythmic stimulus was checked to confirm that the average IOI was at the desired rate and that the overall duration and number of sound elements matched the corresponding isochronous stimulus. If the arrhythmic sequence did not meet these criteria, the process was repeated. In addition, to ensure that the arrhythmic pattern was significantly different from isochrony, a minimum SD was specified. This value was adjusted based on tempo (between 75 and 275 ms) so that each pair of isochronous and arrhythmic stimuli was matched in the number of sound elements, mean tempo, and overall duration ($SD \geq 10$ ms for IOIs ≥ 90 ms; $SD \geq 1$ ms for IOIs between 77 and 89 ms; $SD \geq 0.01$ ms for IOIs < 77 ms; *SI Appendix, Fig. S3A*). To minimize abrupt transitions between sound elements, silent gaps were filled with white noise that had the same spectral energy as quiet periods in the original recordings. For each sequence, the amplitude of each element was constant, but there was some variation in amplitude between sequences made from song elements taken from different recordings (*SI Appendix, Fig. S3B*; note that this amplitude variation was irrelevant to the rhythm discrimination task).

For sounds A, B, C, and E, stimuli were generated at two base tempi: 120 and 180 ms IOI. These tempi were chosen based on the average syllable rate in zebra finch song [~ 7 to 9 syllables per second, or 111 to 142 ms IOI (18)]. For each sound type, a 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 each sound type, the temporal pattern of the arrhythmic probe stimulus was unique. For rule training (see below, *Rule training*), an isochronous/arrhythmic pair was generated with sound D at tempi ranging from 75 to 275 ms IOI in 1 ms steps (i.e., at 201 rates). Once again, each arrhythmic stimulus had a temporally unique pattern. During the generation of these stimuli, sequences at a few rates (15 out of 201) were inadvertently made with higher amplitudes (*SI Appendix, Fig. S3B*).

Auditory Operant Training Procedure. Training and probe testing used a go/interrupt paradigm developed by Lim et al. (64) in which birds had access to a trial switch and a response switch with a water reward spout in the response switch (Fig. 1A). In these experiments, zebra finches were mildly water restricted and worked for water (~ 5 to $10 \mu\text{L}/\text{drop}$), routinely performing ~ 520 trials/day on average. A green light on the trial switch indicated that a trial could be initiated. Pecking the trial switch triggered the playback of a stimulus and extinguished the green light. For each trial, the stimulus had a 50% chance of being a rewarded (S+) or unrewarded stimulus (S−). For all

rhythm discrimination experiments, the S+ stimulus was the isochronous pattern, and the S− stimulus was an arrhythmic pattern. Trial and response switches and their associated lights were activated 500 ms after the stimulus onset, and pecking either switch after this time halted playback. *SI Appendix, Table S1* lists the possible responses and their consequences. "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 to 20 s) was adjusted to control for response bias: the lights-out period was longer the stronger a bird's tendency to respond to every stimulus (65). If neither switch was pecked within 5 s of the 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. Pecking the trial switch was counted as a "miss" or "correct rejection" depending on whether a S+ or S− stimulus had been presented. Birds had to wait 100 ms before the next trial could be initiated.

Shaping and performance criteria. Prior to rhythmic discrimination training, birds were pretrained on the operant conditioning procedure using conspecific songs ("shaping" phase; *SI Appendix, Supplementary Information Text*). Briefly, birds first learned the task of distinguishing between two unfamiliar conspecific songs (~ 2.4 s long), 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 correct responses to the S+ stimulus. In all shaping and training phases, the criterion for advancement to the next phase was $\geq 60\%$ hits, $\geq 60\%$ correct rejections, and $\geq 75\%$ overall correct for two of three consecutive days. Two males did not complete the shaping process within 30 d 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, he was trained to discriminate between isochronous and arrhythmic stimuli ($n = 10$ birds). To minimize the possibility of overlearning a particular tempo or fixating on an acoustic feature other than temporal regularity, each subject was trained using multiple sound types (cf. Fig. 1B) and multiple stimulus rates, or "tempi" (Fig. 1C). In the first phase of training, each bird learned to discriminate two isochronous stimuli (120 and 180 ms IOI) from two arrhythmic stimuli matched for mean IOI. Once that discrimination was learned, a new set of stimuli at the same tempo but with a novel sound element were introduced (Fig. 1D). The subjects learned sounds in one of two orders: one group of males ("ABC"; $n = 4$) learned to discriminate sound A stimuli followed by sound B and then sound C; a second group ("CAB"; $n = 6$) started with sound C followed by sounds A and B. One ABC bird was presented with stimuli at three additional tempi (137.5, 150, and 157.5 ms IOI), but those trials were not reinforced and were excluded from analysis. Upon reaching the criteria for discrimination for training stimuli, the reinforcement rate was reduced from 100 to 80% for at least 2 d prior to "probe testing" (see below, *Probe testing/generalization*) to acclimate the subjects to an occasional lack of reinforcement. More detailed methods of the operant chamber and training procedure are described in *SI Appendix, Supplementary Information Text*.

Probe testing/generalization. To assess whether birds could generalize the isochronous versus arrhythmic classification to a novel (untrained) tempo, they were tested with probe stimuli at a tempo of 144 ms IOI. Pilot testing showed that birds performed better on probe tests after training with multiple stimulus sets. Therefore, 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; Fig. 1D). 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. A total of 10% of the interleaved training stimuli also were not reinforced or punished so that the lack of reinforcement was not unique to probe stimuli (24).

Rule training. After probe testing with three sound types (Fig. 1D), subjects moved on to a training phase with a broader stimulus set using a novel sound (sound D; Fig. 1B). This stimulus set included every integer rate between 75 to 275 ms IOI for a total of 201 rates or 402 stimuli (one isochronous and one arrhythmic stimulus per tempo). Each arrhythmic stimulus ($n = 201$ in total) was generated independently, making it less likely that performance depended on a particular arrhythmic pattern. Every trial was randomly drawn from these stimuli with replacement. All trials were reinforced with water or punished with lights out as during training, but the large number of stimuli and variability in rate across trials made it unlikely that a bird could memorize all of the individual stimuli.

To investigate whether an exposure to a wide range of rates and patterns enhanced or solidified the regularity discrimination, a subset of birds that completed rule training ($n = 5$ of 7) were subsequently trained and tested with an additional stimulus set (sound E).

Tempo discrimination. To determine whether male zebra finches can discriminate between two stimuli that differ in tempo by 20%, we tested a separate cohort of males on tempo discrimination ($n = 4$; 72.8 ± 6.6 [SD] dph at start of shaping). Following shaping, these birds were trained using isochronous sequences of sound A at two rates: 120 and 144 ms IOI. The subjects were counterbalanced so that the rewarded (S+) stimulus was 120 ms IOI for two birds and 144 ms IOI for the other two. Learning curves for tempo training are shown in *SI Appendix, Fig. S1*. The proportion correct was computed based on the last 500 trials.

Data Analysis.

Training and generalization testing. To quantify overall performance and performance during 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). This proportion was then compared to performance at chance ($P = 0.5$) with a binomial test using $\alpha = 0.05/3$ tempi in the testing conditions and $\alpha = 0.05/2$ tempi in the training conditions (Bonferroni correction).

The amount the trial switch was used to interrupt trials varied widely among birds and therefore was not analyzed further in this study.

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 hits and all false alarms for each stimulus within a phase. For rhythmic

discrimination training, reaction times were averaged over 1,500 trials (last 500 trials for each training phase in Fig. 1D).

Rule training. To minimize any potential effect of memorization, we analyzed the first 1,000 trials for each bird. Trials were binned in 10 ms increments starting at 75 ms IOI (20 bins), and the number of correct responses were 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, using the lme4 (*glmer*) statistical package for R (version 3.6.2) within RStudio (version 1.2.5033). Performance in each bin was compared to performance during the 75 to 85 ms bin in which each subject's performance fell to chance because of stimulus generation constraints.

Data Availability. Concatenated trial data, stimulus files, circuit diagrams, analysis code, operant chamber setup diagram, and summarized data files are available through Mendeley Data: <https://dx.doi.org/10.17632/fw5f2vrf4k.1>. 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 also thank R. Dooling and his laboratory for discussion and pilot testing of tempo discrimination thresholds in adult zebra finches. We thank C. ten Cate for helpful discussions on the training procedure, analysis, and interpretation of results and D. Barch and J. M. Reed for helpful discussions on the statistical analyses. 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.).

1. S. Schöneich, K. Kostarakos, B. Hedwig, An auditory feature detection circuit for sound pattern recognition. *Sci. Adv.* **1**, e1500325. (2015).
2. S. A. Kotz, A. Ravignani, W. T. Fitch, The evolution of rhythm processing. *Trends Cogn. Sci.* **22**, 896–910 (2018).
3. J. A. Grahm, Neural mechanisms of rhythm perception: Current findings and future perspectives. *Top. Cogn. Sci.* **4**, 585–606 (2012).
4. B. Morillon, S. Baillet, Motor origin of temporal predictions in auditory attention. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E8913–E8921 (2017).
5. J. J. Cannon, A. D. Patel, How beat perception coopts motor neurophysiology. *Trends Cogn. Sci.* **25**, 135–150 (2021).
6. J. A. Grahm, M. Brett, Rhythm and beat perception in motor areas of the brain. *J. Cogn. Neurosci.* **19**, 893–906 (2007).
7. S. J. Kung, J. L. Chen, R. J. Zatorre, V. B. Penhune, Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat. *J. Cogn. Neurosci.* **25**, 401–420 (2013).
8. S. C. Vernes, G. S. Wilkinson, Behaviour, biology and evolution of vocal learning in bats. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190061 (2020).
9. C. ten Cate, M. Spierings, J. Hubert, H. Honing, Can birds perceive rhythmic patterns? A review and experiments on a songbird and a parrot species. *Front. Psychol.* **7**, 730 (2016).
10. P. Norton, C. Scharff, “Bird song metronomics”: Isochronous organization of zebra finch song rhythm. *Front. Neurosci.* **10**, 309 (2016).
11. T. C. Roeske, O. Tchernichovski, D. Poeppel, N. Jacoby, Categorical rhythms are shared between songbirds and humans. *Curr. Biol.* **30**, 3544–3555.e6 (2020).
12. E. D. Jarvis, Evolution of vocal learning and spoken language. *Science* **366**, 50–54 (2019).
13. A. Celma-Miralles, J. M. Toro, Discrimination of temporal regularity in rats (*Rattus norvegicus*) and humans (*Homo sapiens*). *J. Comp. Psychol.* **134**, 3–10 (2020).
14. A. Ravignani, G. Madison, The paradox of isochrony in the evolution of human rhythm. *Front. Psychol.* **8**, 1820 (2017).
15. C. E. Hagmann, R. G. Cook, Testing meter, rhythm, and tempo discriminations in pigeons. *Behav. Processes* **85**, 99–110 (2010).
16. S. H. Hulse, J. Humpal, J. Cynx, Discrimination and generalization of rhythmic and arrhythmic sound patterns by European starlings (*Sturnus vulgaris*). *Music Percept.* **1**, 442–464 (1984).
17. C. ten Cate, M. Spierings, Rules, rhythm and grouping: Auditory pattern perception by birds. *Anim. Behav.* **151**, 249–257 (2019).
18. R. A. Zann, *The Zebra Finch: A Synthesis of Field and Laboratory Studies* (Oxford University Press, 1996).
19. E. Akutagawa, M. Konishi, New brain pathways found in the vocal control system of a songbird. *J. Comp. Neurol.* **518**, 3086–3100 (2010).
20. J. Lampen, K. Jones, J. D. McAuley, S. E. Chang, J. Wade, Arrhythmic song exposure increases ZENK expression in auditory cortical areas and nucleus taeniae of the adult zebra finch. *PLoS One* **9**, e108841 (2014).
21. S. Ma, A. Ter Maat, M. Gahr, Neurotelemetry reveals putative predictive activity in HVC during call-based vocal communications in zebra finches. *J. Neurosci.* **40**, 6219–6227 (2020).
22. J. I. Benichov *et al.*, The forebrain song system mediates predictive call timing in female and male zebra finches. *Curr. Biol.* **26**, 309–318 (2016).
23. J. I. Benichov, D. Vallentin, Inhibition within a premotor circuit controls the timing of vocal turn-taking in zebra finches. *Nat. Commun.* **11**, 221 (2020).
24. J. van der Aa, H. Honing, C. ten Cate, The perception of regularity in an isochronous stimulus in zebra finches (*Taeniopygia guttata*) and humans. *Behav. Processes* **115**, 37–45 (2015).
25. M. A. Long, M. S. Fee, Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* **456**, 189–194 (2008).
26. S. J. Spaulding *et al.*, Cueing and gait improvement among people with Parkinson's disease: A meta-analysis. *Arch. Phys. Med. Rehabil.* **94**, 562–570 (2013).
27. S. E. Chang, F. H. Guenther, Involvement of the cortico-basal ganglia-thalamocortical loop in developmental stuttering. *Front. Psychol.* **10**, 3088 (2020).
28. C.-E. Benoit *et al.*, Musically cued gait-training improves both perceptual and motor timing in Parkinson's disease. *Front. Hum. Neurosci.* **8**, 494 (2014).
29. E. Ladányi, V. Persici, A. Fiveash, B. Tillmann, R. L. Gordon, Is atypical rhythm a risk factor for developmental speech and language disorders? *Wiley Interdiscip. Rev. Cogn. Sci.* **11**, e1528 (2020).
30. S. E. Trehub, L. A. Thorpe, Infants' perception of rhythm: Categorization of auditory sequences by temporal structure. *Can. J. Psychol.* **43**, 217–229 (1989).
31. H. Merchant, J. Grahm, L. Trainor, M. Rohrmeier, W. T. Fitch, Finding the beat: A neural perspective across humans and non-human primates. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20140093 (2015).
32. T. E. Matthews, M. A. G. Witek, T. Lund, P. Vuust, V. B. Penhune, The sensation of groove engages motor and reward networks. *Neuroimage* **214**, 116768 (2020).
33. J. T. Coull, R.-K. Cheng, W. H. Meck, Neuroanatomical and neurochemical substrates of timing. *Neuropsychopharmacology* **36**, 3–25 (2011).
34. D. L. Bowling, D. Purves, A biological rationale for musical consonance. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11155–11160 (2015).
35. A. Bottcher, S. Gerö, K. Beedholm, H. Whitehead, P. T. Madsen, Variability of the inter-pulse interval in sperm whale clicks with implications for size estimation and individual identification. *J. Acoust. Soc. Am.* **144**, 365–374 (2018).
36. J. N. Schneider, E. Mercado, Characterizing the rhythm and tempo of sound production by singing whales. *Bioacoustics* **28**, 239–256 (2019).
37. K. I. Nagel, H. M. McLendon, A. J. Doupe, Differential influence of frequency, timing, and intensity cues in a complex acoustic categorization task. *J. Neurophysiol.* **104**, 1426–1437 (2010).
38. B. Samuels, J. Grahm, M. J. Henry, S. A. MacDougall-Shackleton, European starlings (*Sturnus vulgaris*) discriminate rhythms by rate, not temporal patterns. *J. Acoust. Soc. Am.* **149**, 2546 (2021).
39. F. L. Bouwer, V. Nityananda, A. A. Rouse, C. ten Cate, Rhythmic abilities in humans and non-human animals: A review and recommendations from a methodological perspective. *Philos. Trans. R. Soc. B Biol. Sci.*, 10.1098/rstb.2020.0335 Published ahead of print.
40. M. Wirthlin *et al.*, A modular approach to vocal learning: Disentangling the diversity of a complex behavioral trait. *Neuron* **104**, 87–99 (2019).
41. C. Scharff, M. Knörnschild, E. D. Jarvis, “Vocal learning and spoken language: Insights from animal models with an emphasis on genetic contributions” in *Human Language: From Genes and Brains to Behavior*, P. Hagoort, Ed. (MIT Press, 2019), pp. 657–685.
42. R. Mooney, The neurobiology of innate and learned vocalizations in rodents and songbirds. *Curr. Opin. Neurobiol.* **64**, 24–31 (2020).
43. Y. S. Zhang, A. A. Ghazanfar, A hierarchy of autonomous systems for vocal production. *Trends Neurosci.* **43**, 115–126 (2020).

44. D. E. Okobi Jr, A. Banerjee, A. M. M. Matheson, S. M. Phelps, M. A. Long, Motor cortical control of vocal interaction in neotropical singing mice. *Science* **363**, 983–988 (2019).
45. A. Banerjee, S. M. Phelps, M. A. Long, Singing mice. *Curr. Biol.* **29**, R190–R191 (2019).
46. N. Katsu, K. Yamada, K. Okanoya, M. Nakamichi, Temporal adjustment of short calls according to a partner during vocal turn-taking in Japanese macaques. *Curr. Zool.* **65**, 99–105 (2019).
47. D. Y. Takahashi, A. R. Fenley, A. A. Ghazanfar, Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20150370 (2016).
48. L. S. Burchardt, P. Norton, O. Behr, C. Scharff, M. Knörnschild, General isochronous rhythm in echolocation calls and social vocalizations of the bat *Saccopteryx bilineata*. *R. Soc. Open Sci.* **6**, 181076 (2019).
49. D. Genzel, J. Desai, E. Paras, M. M. Yartsev, Long-term and persistent vocal plasticity in adult bats. *Nat. Commun.* **10**, 3372 (2019).
50. A. Ravignani et al., What pinnipeds have to say about human speech, music, and the evolution of rhythm. *Front. Neurosci.* **10**, 274 (2016).
51. C. Reichmuth, C. Casey, Vocal learning in seals, sea lions, and walruses. *Curr. Opin. Neurobiol.* **28**, 66–71 (2014).
52. K. Ralls, P. Fiorelli, S. Gish, Vocalizations and vocal mimicry in captive harbor seals, *Phoca vitulina*. *Can. J. Zool.* **63**, 1050–1056 (1985).
53. A. L. Stansbury, V. M. Janik, Formant modification through vocal production learning in gray seals. *Curr. Biol.* **29**, 2244–2249.e4 (2019).
54. P. Cook, A. Rouse, M. Wilson, C. Reichmuth, A California sea lion (*Zalophus californianus*) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *J. Comp. Psychol.* **127**, 412–427 (2013).
55. A. A. Rouse, P. F. Cook, E. W. Large, C. Reichmuth, Beat keeping in a sea lion as coupled oscillation: Implications for comparative understanding of human rhythm. *Front. Neurosci.* **10**, 257 (2016).
56. M. D. Beecher, E. A. Brenowitz, Functional aspects of song learning in songbirds. *Trends Ecol. Evol.* **20**, 143–149 (2005).
57. F. Nottebohm, M. E. Nottebohm, L. Crane, Developmental and seasonal changes in canary song and their relation to changes in the anatomy of song-control nuclei. *Behav. Neural Biol.* **46**, 445–471 (1986).
58. T. Noda, T. Amemiya, T. I. Shiramatsu, H. Takahashi, Stimulus phase locking of cortical oscillations for rhythmic tone sequences in rats. *Front. Neural Circuits* **11**, 2 (2017).
59. M. M. Asokan, R. S. Williamson, K. E. Hancock, D. B. Polley, Inverted central auditory hierarchies for encoding local intervals and global temporal patterns. *Curr. Biol.* **31**, 1762–1770.e4 (2021).
60. F. L. Bouwer, H. Honing, H. A. Slagter, Beat-based and memory-based temporal expectations in rhythm: Similar perceptual effects, different underlying mechanisms. *J. Cogn. Neurosci.* **32**, 1221–1241 (2020).
61. M. Espinoza-Monroy, V. de Lafuente, Discrimination of regular and irregular rhythms explained by a time difference accumulation model. *Neuroscience* **459**, 16–26 (2021).
62. T. F. Roberts et al., Identification of a motor-to-auditory pathway important for vocal learning. *Nat. Neurosci.* **20**, 978–986 (2017).
63. H. Honing, F. L. Bouwer, L. Prado, H. Merchant, Rhesus monkeys (*Macaca mulatta*) sense isochrony in rhythm, but not the beat: Additional support for the gradual audiomotor evolution hypothesis. *Front. Neurosci.* **12**, 475 (2018).
64. Y. Lim, R. Lagoy, B. G. Shinn-Cunningham, T. J. Gardner, Transformation of temporal sequences in the zebra finch auditory system. *eLife* **5**, 1–18 (2016).
65. A. Gess, D. M. Schneider, A. Vyas, S. M. N. Woolley, Automated auditory recognition training and testing. *Anim. Behav.* **82**, 285–293 (2011).