

The motor origins of human and avian song structure

Adam T. Tierney^{a,b}, Frank A. Russo^c, and Aniruddh D. Patel^{b,1}

^aAuditory Neuroscience Laboratory, Northwestern University, Evanston, IL 60208; ^bThe Neurosciences Institute, San Diego, CA 92121; and ^cDepartment of Psychology, Ryerson University, Toronto, ON, Canada M5B 2K3

Edited by Dale Purves, Duke University Medical Center, Durham, NC, and approved July 20, 2011 (received for review March 30, 2011)

Human song exhibits great structural diversity, yet certain aspects of melodic shape (how pitch is patterned over time) are widespread. These include a predominance of arch-shaped and descending melodic contours in musical phrases, a tendency for phrase-final notes to be relatively long, and a bias toward small pitch movements between adjacent notes in a melody [Huron D (2006) *Sweet Anticipation: Music and the Psychology of Expectation* (MIT Press, Cambridge, MA)]. What is the origin of these features? We hypothesize that they stem from motor constraints on song production (i.e., the energetic efficiency of their underlying motor actions) rather than being innately specified. One prediction of this hypothesis is that any animals subject to similar motor constraints on song will exhibit similar melodic shapes, no matter how distantly related those animals are to humans. Conversely, animals who do not share similar motor constraints on song will not exhibit convergent melodic shapes. Birds provide an ideal case for testing these predictions, because their peripheral mechanisms of song production have both notable similarities and differences from human vocal mechanisms [Riede T, Goller F (2010) *Brain Lang* 115:69–80]. We use these similarities and differences to make specific predictions about shared and distinct features of human and avian song structure and find that these predictions are confirmed by empirical analysis of diverse human and avian song samples.

birdsong | evolution | music

Song exists in every human culture and exhibits a vast diversity of forms (1). What regularities exist amid this diversity, and what is their origin? This question is relevant to debates over the evolutionary and biological foundations of music (2, 3). One widespread feature is a tendency to create melodies from a limited set of stable pitches or pitch intervals (i.e., the use of musical “scales”) (4, 5). This may reflect a human cognitive tendency to create sound sequences from a small inventory of discrete elements, which are generatively recombined to form unique patterns. [This propensity is well known from language, in which unique sentences are produced from a finite set of phonemes (5).] Although musical scales and pitch intervals have recently attracted attention from biologists (6, 7), much less attention has been paid to the melodic “shape” of human song (i.e., how pitch patterns unfold over time). Cross-cultural research has revealed several widespread aspects of melodic shape, including (i) a predominance of arch-shaped and descending melodic contours in musical phrases, (ii) a tendency for phrase-final notes to be relatively long, and (iii) a bias toward “smooth” pitch contours (i.e., a statistical bias toward small pitch movements between adjacent pitches in a melody) (8–10) (Fig. 1).

What is the origin of these features? In the study of speech, universal design principles are often taken as the hallmark of biological specialization for language (11). In contrast, we hypothesize that many widespread features in the melodic shape of song stem from motor constraints on song production, rather than being innately specified. By “motor constraints” we mean that the sound-producing actions underlying different melodic shapes vary in their energetic cost and that less costly actions are favored (cf. 12). Consequently, certain melodic shapes are widespread. A brief analogy helps to illustrate this argument: when humans swim they can adopt many different motor patterns, including the “butterfly” (which involves simultaneously lifting both arms out of

the water and rotating them about the shoulder) and the “crawl” (which involves lifting one arm out of the water at a time, while rolling the body to the side). Both are effective ways of swimming, but because of the biomechanics of the human body in water, the crawl is less energetically expensive than the butterfly. According to a motor constraint hypothesis, this difference accounts for the fact that the crawl is a much more commonly observed swimming pattern than the butterfly.

The motor constraint hypothesis for human song claims that energetic costs (i.e., the metabolic energy required for production) underlie widespread features of melodic shape described above and leads to testable predictions. One prediction of this hypothesis is that any animals subject to similar motor constraints on song will exhibit similar melodic shapes, no matter how distantly related those animals are to humans. Conversely, animals who do not share the same motor constraints on song will not exhibit convergent melodic shapes. Birds provide an ideal case for testing these predictions. The ancestors of humans and birds diverged more than 250 million years ago, and the functional anatomy of humans and birds differs in many respects (13). Nevertheless, birdsong specialists have recently emphasized several (convergently evolved) commonalities in the vocal production biomechanics of birds and humans, while also noting specific differences (14). Two major commonalities are (i) birds and humans use respiratory air pressure to drive sound-producing oscillations in membranous tissues (the vocal folds in humans; the labial folds in birds), and (ii) the resulting sounds are filtered by a vocal tract whose shape can be rapidly changed to emphasize or attenuate certain frequencies (15–20). A major difference between birds and humans concerns the fact that birds have two sets of oscillating membranes that can be controlled independently (in the two-sided syrinx), whereas humans only have one set (in the larynx) (14).

On the basis of our motor constraint hypothesis, we use these biomechanical similarities and differences to predict specific acoustic similarities and differences in the melodic shape of human and avian song, as detailed below. Testing these predictions requires empirical comparison of human and avian song structure based on diverse samples in both domains. Hence the selection of materials for analysis was an important part of this study. For human data we chose folk songs, using the largest cross-cultural database of digitally encoded folk songs that have been segmented into individual phrases: the Essen Musical database (21). This corpus contains music notation for 9,467 folk songs (52,899 phrases) from 32 geographic regions including Austria, China, France, Germany, Ireland, Luxembourg, The Netherlands, Nova Scotia, Russia, Scotland, Switzerland, and Yugoslavia (Table S1). Although this represents a small sample of the world’s cultures, the inclusion of many songs from China (>2,000) allows us to test whether the patterns we study are primarily European or are more general in nature. Of course, any widespread song features in this

Author contributions: A.T.T., F.A.R., and A.D.P. designed research; A.T.T. and A.D.P. performed research; A.T.T. and F.A.R. analyzed data; and A.T.T. and A.D.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be sent. E-mail: apatel@nsi.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1103882108/-DCSupplemental.

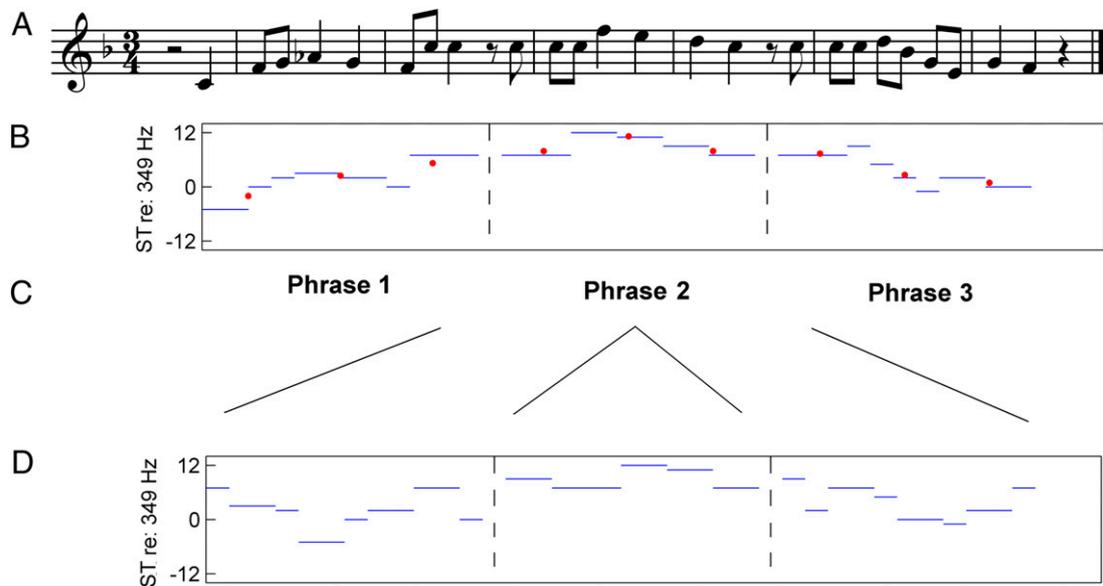


Fig. 1. Example human melody, illustrating arch-shaped and descending melodic contours, relatively long phrase-final notes, and smooth pitch contours. (A) A German folk song from the Essen database, in Western music notation (Audio File S1). The song consists of three phrases. (B) The same song represented as a pattern of pitch vs. time in semitones (ST) from the tonic pitch (F4 or 349 Hz, the last pitch of the melody). The blue horizontal lines show the pitch values of individual tones and their relative durations, and phrase boundaries are indicated by dashed vertical lines (in this song, all tones have one of two possible durations, short and long: note how phrase-final tones are always long). Red dots within each phrase show the mean pitch of the first, second, and last third of each phrase. (C) Melodic contour shapes (ascending, arch, and descending) assigned to the three phrases according to the pitch patterns of the red dots in B (see text for details). (D) The melody in B with the individual notes in each phrase randomly reordered in time, illustrating the more jagged contours that result (i.e., a tendency for larger jumps between adjacent pitches, compared with the original melody, which has a relatively “smooth” pitch contour by comparison). In the Essen database, phrases had an average of 9.5 notes each, excluding rests ($SD = 3.4$) and an average range of 8.9 semitones (st) between the highest and lowest note of the phrase ($SD = 3.7$). Songs had an average of 5.6 phrases each ($SD = 2.8$).

database cannot be claimed as “universals,” yet they merit attention because the motor constraint hypothesis predicts that these patterns will be found across the diversity of human cultures. A sample folk song from this database is shown in Fig. 1 A and B.

For bird data we chose songs in which all notes had either a pure-tone quality or a harmonic structure with a fundamental frequency as the harmonic with the most power. That is, we focused on “tonal” birdsongs, to conduct empirical comparisons of pitch patterns in avian and human song. In addition, we focused on birdsongs with at least five notes, low background noise, and significant pitch variation (*Experimental Methods*). Using multiple sources, including research libraries and published recordings, we compiled a taxonomically diverse sample of birdsongs from 54 songbird families, with 80 species represented (one song per species; Table S2). Three birdsongs from our database are shown in Fig. 2. In gathering birdsongs we focused on songbird (oscine) families because they are the richest source of tonal songs in the avian world. Like humans, oscines are vocal learners. Some suboscines also produce songs, but these songs are not learned from an auditory model (22). Although we focus on songbirds in this study, our motor constraint hypothesis should also apply to suboscines that produce tonal songs with peripheral vocal mechanisms akin to those in songbirds.

Our empirical comparative analyses followed the following strategy: for each of the three widespread aspects of melodic shape under investigation, we examined our human and birdsong corpora, first confirming its existence in human song and then testing for specific similarities and differences to birdsong, on the basis of the predictions of the motor constraint hypothesis. In conducting this work, it is important to consider whether our selection criteria for birdsongs (i.e., tonal songs with significant pitch variation) may have biased us toward finding certain kinds of pitch patterns. Our criteria excluded birdsongs containing

a more varied selection of sounds, including buzzy, noisy, or click-like notes containing inharmonic frequencies (e.g., the songs of European starlings, *Sturnus vulgaris*). Instead, we focus on birdsongs with notes containing strong fundamental frequency contours, resulting in clear pitch patterns. Crucially, however, the mere existence of clear pitch patterns in birdsongs does not, a priori, imply that these patterns will follow any particular shape. Thus, we believe the outcome of our study is not predetermined by our selection criteria.

One other possible concern with our approach is that our analysis of human song relies on notation rather than acoustic recordings, whereas our analysis of birdsong focuses on recordings. Ideally, our analyses would focus on recordings in both domains, but extensive audio corpora of culturally diverse human folk songs, consisting of single melodic lines segmented into individual notes and phrases, are not readily available. Thus, our analysis of notation raises the question of how closely the patterns we observe reflect those found in actual human singing. This is an important question, because it is known that pitch and timing patterns in music performance are not identical to the values indicated by notation (23). For example, a singer may “bend” (lower or raise) the pitch of a note, or alter a note’s duration, for expressive effect. However, our analyses of pitch patterns do not focus on the fine-grained nuances of melodic shape but on broad features such as overall contour and the average size of pitch intervals in a melody, which are similar in performance and notation (24). Furthermore, our analysis of phrase-final lengthening is conservative, because research shows that in music performance the final notes of phrases are even longer than indicated by notation (25, 26). Thus, we fully expect that our notation-based findings will be replicated when broad-based audio corpora of human song are analyzed in the future.

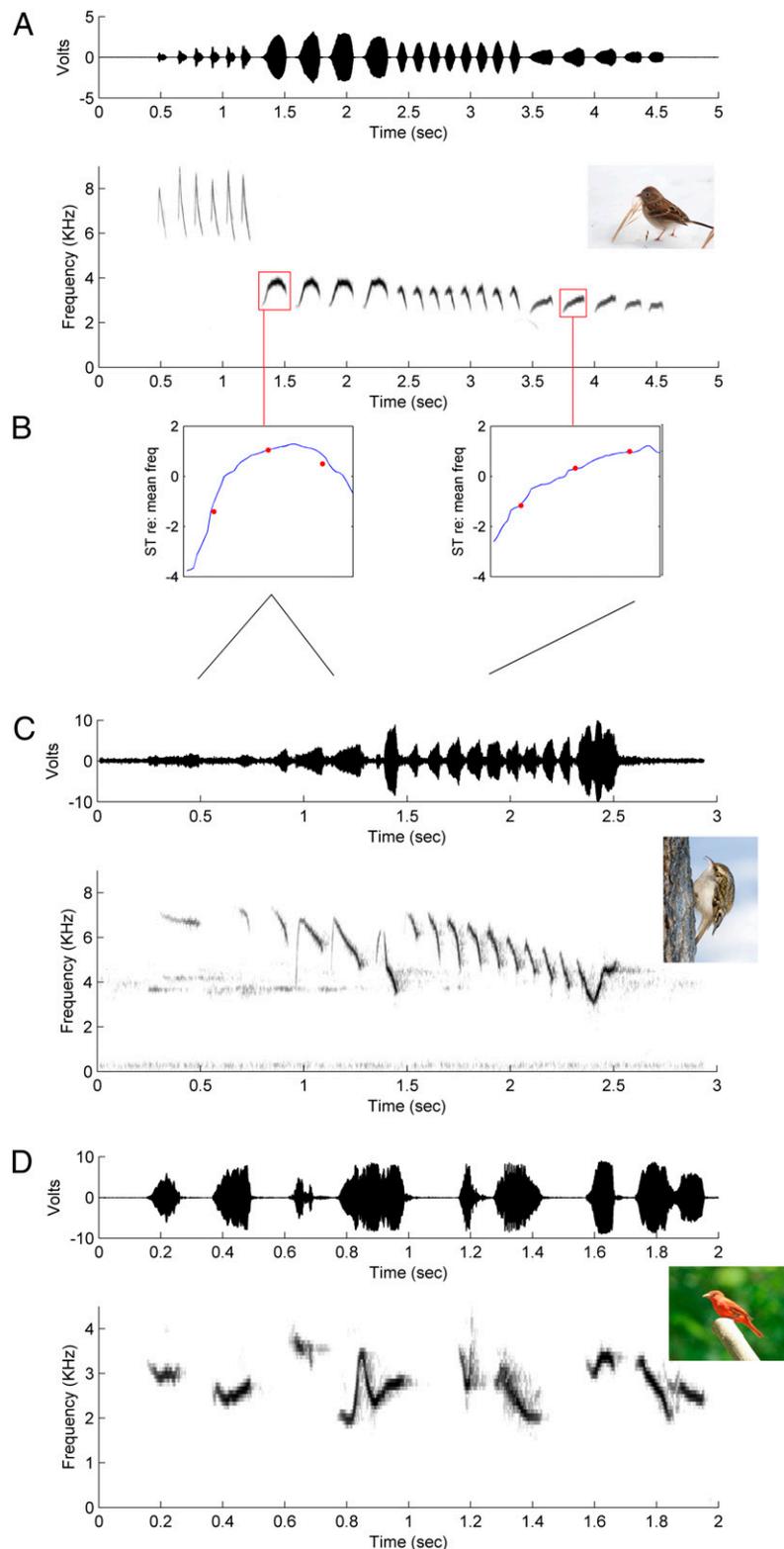


Fig. 2. Example bird songs (Audio Files S2, S3, S4, S5, and S6). (A) Waveform and spectrogram of a field sparrow song, *Spizella pusilla* (family Emberizidae), to illustrate the pitch contours of individual notes. (B) Two notes from the birdsong represented as pitch-time contours, with the mean pitches of each third of the note shown by red dots, as in Fig. 1B. Shown below the pitch contours are the melodic contour shapes assigned to these two notes. (Shape classification was done for all birdsong notes: the notes in B are shown for illustrative purposes. Several more examples of birdsong note shape classification are given in Figs. S3–S7.) (C) Waveform and spectrogram of a Eurasian treecreeper song, *Certhia familiaris* (family Certhiidae), to illustrate the tendency for long final notes. (D) Waveform and spectrogram of a summer tanager, *Piranga rubra* (family Thraupidae), to illustrate a tendency for large pitch jumps between adjacent notes. Birdsongs in our corpus averaged 2 s in duration ($SD = 1.2$), had 13.7 notes on average ($SD = 8.1$), and an average range of 9.0 semitones between the median frequencies of the highest and lowest notes of the song ($SD = 4.5$). On the basis of the median frequency of each note in the corpus, the average frequency of birdsong notes was 3,730 Hz ($SD = 1,584$). Field sparrow image courtesy of Kelly Colgan Azar. Eurasian treecreeper image courtesy of Sergey Yeliseev. Summer tanager image courtesy of Jeff Whitlock.

Results

Melodic Contour and Respiratory Constraints. Human song phrases, like spoken utterances, are produced during exhalations in which air pressure beneath the vocal folds (“subglottal pressure”) is regulated to influence the loudness and pitch of sound (26). Subglottal pressure during these exhalations has a characteristic profile: it rises rapidly near the beginning of the utterance, stays relatively steady or declines slightly during most of the utterance (with smaller modulations involved in regulating loudness and pitch), and falls sharply at the end (27). Humans can control the pitch of their voice independently of subglottal pressure (via the tension of the vocal folds), yet other things being equal, higher pressure leads to faster vocal fold vibration and hence higher pitch (28). Thus, higher pitches should be easier to produce when subglottal pressure is high, and vice versa. On the basis of this motor constraint one would expect two types of pitch contours to predominate in human song phrases: arch-shaped contours that rise to a peak and then fall, and descending contours that start high and gradually lower in pitch over the course of a phrase.

Following the method of Huron (9), we classified all song phrases in our human corpus into one of nine melodic contour shapes: rising, falling, rising-falling (arch), falling-rising, etc. Specifically, each phrase was converted to a pattern of pitch vs. time, divided into three equal time segments, and the mean pitch of each segment was taken. The resulting three pitch values formed a pattern that was classified into one of nine possible shapes, depending on whether the first and last pitches were higher, lower, or equivalent to the middle pitch. Fig. 3A shows a histogram of the resulting shapes. Confirming Huron’s earlier findings (based on European folk songs), arch and descending contours were the most common melodic shapes. Notably, arch contours were significantly more common than their inverse shape (V-shape), and descending contours were significantly more common than their inverse ascending shape (both $P < 0.0001$, binomial test). (These patterns held for the corpus as a whole and for Chinese songs analyzed separately.) These distributional biases are predicted by the motor-constraint hypothesis, because the biomechanical relationship between subglottal pressure and vocal fold vibration rate should make arch and descending pitch contours more energetically efficient to produce than their inverse shapes.

Songbird vocalizations are primarily produced during controlled exhalation, and given the similarities in the myoelastic-

aerodynamic sound-producing mechanisms in humans and birds (14), the motor constraint hypothesis predicts that arch and descending contours will predominate in birdsong. In testing this prediction, one important difference between birds and humans was taken into account. Unlike humans, birds tend to breathe between the individual notes of their songs, likely because this allows them to sing longer songs by constantly replenishing the small air sacs that supply the lungs (17). This means that each note is produced by a separate small exhalation, which leads to the prediction that arch and descending contours will predominate at the level of *individual notes*. To test this prediction we used computer software to track the fundamental frequency contours of all notes in our corpus ($n = 1,092$) on sound spectrograms (*Experimental Methods*). We then converted each frequency contour into a pitch contour and classified its shape using methods identical to those described above for human song phrases (Fig. 2B; note that conversion of avian contours from a linear Hz scale to a logarithmic semitone scale did not influence the pattern of results; *Experimental Methods*). Fig. 3B shows the resulting distribution. Just as with human song, arch and descending contours were common shapes. Furthermore, arches were significantly more common than V-shaped contours, and descending contours were significantly more common than ascending contours, as predicted by the motor constraint hypothesis (both $P < 0.0001$, binomial test).

The analysis above depended on reducing each pitch–time contour to three pitch values, to classify them into different shapes. This procedure suggested that arch and declining contours were dominant shapes in both human and avian song. As an independent check on this finding, we conducted an ancillary analysis that used much more pitch information from each contour. Specifically, for human song phrases and birdsong notes, we normalized the duration of all pitch–time contours and then averaged them, to produce average melodic shapes in each domain. The average shapes of both human song phrases and birdsong notes reflected the dominance of arch and declining contours, supporting the findings of our main analysis (Fig. S1).

Phrase-Final Note Duration and Articulatory Constraints. In human song (and speech) the last note of phrases tends to be relatively long (25, 26). To examine our corpus for this pattern, we computed the relative duration of all phrase-final notes in our human songs. Specifically, for each phrase we computed each note’s duration relative to the average duration of all notes in that phrase (*Experimental Methods*). Across all phrases in the corpus, we then averaged the relative duration of phrase-final notes and nonfinal notes. Phrase-final notes were significantly longer than nonfinal notes [mean (SD) of 1.58 (0.74) vs. 0.93 (0.39), $t = 134.3$, $P < 0.0001$]. Again, this held for the corpus as a whole and for Chinese songs analyzed separately.

What motor constraint might underlie this pattern? Human song (like speech) is characterized by rapid changes in the shape of the vocal tract, which serve to change its resonating properties (26). These movements typically cease momentarily at song phrase boundaries (e.g., when drawing the next breath), and the ease of slowing the articulators before coming to a complete stop, vs. stopping abruptly, could underlie the tendency for relatively long notes at phrase endings (29). Like humans, many birds also actively and rapidly change the shape of their vocal tract during song production to emphasize or attenuate certain frequencies (15, 30). Because these movements cease at the end of a song, the motor constraint hypothesis predicts that birdsong will tend to show long final notes. For each birdsong we computed the relative duration of each note by dividing each note’s duration by the mean note duration in the song. By averaging the relative duration of song-final notes vs. nonfinal notes, we found that final notes tended to be relatively long, just as in human song [mean

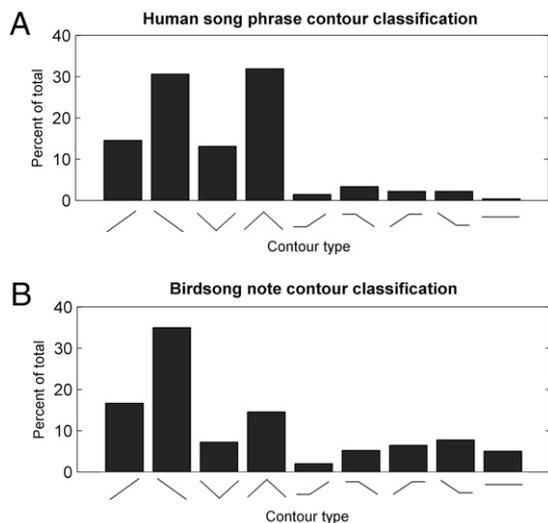


Fig. 3. Distribution of melodic contour shapes in (A) human song phrases and (B) individual birdsong notes. Following ref. 9, all contours were assigned to one of nine possible shapes, as described in the text.

(SD) of 1.33 (0.74) vs. 0.97 (0.55), $t = 7.6$, $P < 0.0001$]. A birdsong with a relatively long final note is shown in Fig. 2C.

Small Pitch Jumps and Vibratory Constraints. It has long been observed that the jumps between adjacent pitches in human songs tend to be small (i.e., there is an overall bias toward small pitch intervals and hence “smooth” melodic contours) (8, 10, 31). One simple way to demonstrate this is to randomly shuffle the order of pitches in a musical phrase. The resulting random phrase usually has larger jumps between adjacent pitches than the original phrase (10) (Fig. 1D). This shows that the melodic shapes of human musical phrases are smoother than one would expect simply on the basis of the distribution of pitches within a phrase. According to the motor constraint hypothesis, this bias toward small pitch intervals is due to the fact that small pitch jumps are easier to produce than large ones, because large jumps require sudden contraction or relaxation in the muscles controlling vocal fold tension. This hypothesis predicts that birds that have two sets of sound-producing labial folds should be less influenced by this constraint. This is because birds, unlike humans, can adjust tension separately in the labia on the two sides of their syrinx. Thus, by maintaining separate tensions on the left and right pair of labia, large pitch jumps can be made by alternating sound production between these structures, without demanding sudden large changes in the tension of labia on either side (18).

To test this idea, we quantified the degree of bias toward small intervals in human song and birdsong. Inspired by von Hippel (10), we created a measure called the “interval compression ratio” (ICR), defined as the mean absolute interval size (in semitones) for a melody with its pitches randomly reordered, divided by the mean absolute interval size for a melody with its pitches in their original order. (For this analysis, birdsongs were first converted to sequences of discrete frequencies, using the median frequency of each birdsong note. Note that this conversion did not involve mapping birdsong notes onto human musical scales, nor did it force the intervals computed between the discrete frequencies into integer values to resemble human music. Instead, pitch intervals in birdsong could take on continuous values, e.g., 2.3 semitones. Further details in *SI Experimental Methods*.)

The larger the ICR, the more biased a melody is toward small intervals. ICR values for human song phrases and birdsongs were positively skewed and hence were compared using non-parametric statistics. The median ICRs for birdsong and human song phrases were 1.21 and 1.46, respectively (Fig. S2 shows the distributions of avian and human ICR values). This difference was statistically significant (Mann-Whitney U test, $U = 1,099,510$, $P < 0.01$), indicating that human songs were more biased toward small pitch intervals, confirming the predictions of the motor constraint hypothesis. A similar result was obtained via a Monte Carlo analysis in which subsets of the human song corpus were chosen for comparison with birdsong, on the basis of matching the number of notes in human song phrases and birdsongs (*SI Experimental Methods*). The larger ICR value for human vs. avian song held for the Chinese songs analyzed separately, when songs with at least six scale tones were analyzed (*SI Results*). A birdsong with relatively large pitch jumps between adjacent notes is shown in Fig. 2D.

It is worth noting that the decision to exclude birdsongs with minimal pitch variation during our selection of 80 birdsongs for this study should not bias our analysis, because the ICR analysis is agnostic to the absolute amount of pitch variation between notes. It simply compares the average absolute interval size when the pitches of a sequence are randomly reordered vs. when they are in their original order. Hence, as long as the notes of a birdsong have nonzero variance among their median pitch values (which is very likely in biological signals such as birdsongs), the ICR analysis should be valid.

Discussion

What governs the melodic shapes of human songs? By comparing human and avian song we provide evidence that motor constraints, rather than innate factors, are the origin of several widespread features in the structure of human song phrases. Of course, once a regularity exists it can be exploited for communicative functions. Phrase-final lengthening, for example, is regulated in human musical performance as a way of marking structural boundaries (32). In the case of birdsong, it would be interesting to study whether metabolically costly features in song (i.e., those features that go against the motor constraints discussed in this article) are particularly attractive to females, who could potentially use these features to assess male vigor (cf. 12).

A motor constraint hypothesis motivates further comparative work with other species, because it makes testable predictions. For example, it predicts that the pitch contours of vocalizations made while inhaling [e.g., certain “ingressive” sounds made by primates (33)] should not show a bias toward arch and declining pitch contours, because they will not have the characteristic pressure profiles associated with exhalation. Similarly, the motor constraint hypothesis predicts that species that sing without rapidly and actively changing the shape of their vocal tract should not show a tendency to lengthen final notes. Many frogs, for example, produce songs without any salient active changes in vocal tract shape (they produce sound by pumping air through the larynx into a sac that distends passively) (34). Hence the motor constraint hypothesis predicts that frog song, unlike human and avian song, will not show long final notes.

The present work extends a long tradition of comparing the structure of birdsong and human music (e.g., 35–38) but is distinguished by applying empirical methods to a diverse sample of birdsongs and human songs, in the context of hypothesis-driven research. We believe that such an approach can be used to discover many other similarities and differences between animal songs and human music.

Experimental Methods

Pitch and duration values for human songs were imported from the Essen database into MATLAB (MathWorks) using custom-written software in Python. In the Essen database, pitch values are coded as scale steps from the tonic or structurally central pitch of the melody, and duration values are coded as multiples of the shortest note in the song. For melodic contour analyses of phrases, pitch values were converted into continuous functions of pitch vs. time (pitch in semitones from the tonic pitch of the song, time relative to shortest note) (Fig. 1B), and rests were eliminated. Only phrases with at least five notes were used (77% of all phrases in the corpus), to ensure that the pitch–time contours had enough material to assign a meaningful shape. These contours were then sampled at 50 equally spaced time points and then divided into consecutive segments of approximately equal duration (17, 16, and 17 points). The mean pitch of each segment was computed, and the resulting three pitch values were classified into one of nine contour shapes as described in the text. For this analysis pitch values were marked as equivalent if the differences between them did not exceed 0.2 semitones.

For the phrase-final duration analysis of human music, note durations in each phrase were first expressed relative to the shortest note in the phrase [e.g., the third phrase of Fig. 1A would have durations of (1, 1, 1, 1, 1, 1, 2, 2)]. The average duration of all notes in the phrase was then computed, and each note’s duration was expressed relative to this average duration. In doing this analysis, phrases ending in rests were excluded, and if a phrase had internal rests, the durations of these rests were excluded from the analysis. [In the preceding example, the average duration would be 1.22, and relative note duration would thus be (0.82 0.82 0.82 0.82 0.82 1.64 1.64)]. These relative duration values were used to compute the average duration of phrase final vs. nonfinal notes across phrases in the corpus ($N = 36,313$ phrases).

Birdsongs were selected according to the criteria stated in the Introduction. Furthermore, all birdsongs consisted of a sequence of notes preceded and followed by a long pause relative to the duration of the notes and were excluded if the song had minimal pitch variation (less than one semitone difference between the highest and lowest notes, according to the

median frequency of each note) or if any of the notes had two simultaneous, distinct pitches (likely made with the “two voice” properties of the syrinx). We sought songs meeting our criteria from the Cornell Laboratory of Ornithology, the Borror Laboratory of Bioacoustics, the British Museum Library, CDs accompanying *Nature's Music, The Singing Life of Birds*, and *Music of the Birds* (39–41), and 12 Internet sources. Aiming for taxonomic diversity, we originally hoped to collect one song for each of the 84 songbird (oscine) families (42). However, because of our strict criteria and the nature of available materials, we found samples for only 54 families. We then sampled one more species from 26 of these families (sampling when possible from the most speciose families), for a total of 80 songs and species (Table S2). Narrow-band spectrograms were made of each song using SIGNAL (Engineering Design; *SI Experimental Methods*). The duration and fundamental frequency of each note were extracted from spectrograms using an automatic spectral contour detection algorithm (*SI Experimental Methods*). The resulting frequency contours were converted to pitch contours [i.e., all frequency points in the contour were converted to semitones from the mean frequency of the note using the formula $ST = 12 \cdot \log_2(F/\text{mean}(F))$, where F is the frequency of a data point in Hz and $\text{mean}(F)$ is the mean frequency of the note]. (Note that this conversion did not involve mapping birdsong notes into the discrete pitches or intervals of human musical scales but simply converted frequency contours in a linear Hz scale into pitch contours in a logarithmic

semitone scale, to study human and avian pitch contours in a comparable way.) Bird note pitch contours were then sampled at 50 equally spaced time points and classified into shapes using methods identical to those for human song phrases. (Note that a significant predominance of arch and descending contours vs. their inverse shapes was found whether frequency contours in Hz or pitch contours in semitones were used as the basis for birdsong analysis.)

For song-final duration analyses of birdsongs, the durations of notes in each song were first expressed relative to the shortest note in the song. Analysis then proceeded in a manner identical to that for human song-phrases.

ACKNOWLEDGMENTS. We thank Rindy Anderson, Tim Gentner, John Iversen, and Susan Peters for helpful comments; Eleanor Selfridge-Field and Craig Sapp for help with the Essen database, including access to the Chinese folk songs; Lola Cuddy, Christopher Sturdy, and Ronald Weisman for their contributions to an early implementation of the motor constraint hypothesis in the context of birdsong; and staff members at the Cornell Laboratory of Ornithology, the Borror Laboratory of Bioacoustics, and the British Museum Library for their help in locating birdsongs for our study. This study was supported by Neurosciences Research Foundation as part of its program on music and the brain at The Neurosciences Institute, where A.D.P. is the Esther J. Burnham Senior Fellow.

- Nettl B, Stone R, eds (1998) *The Garland Encyclopedia of World Music*. (Garland Publications, New York).
- Wallin NL, Merker B, Brown S, eds (2001) *The Origins of Music* (MIT Press, Cambridge, MA).
- Patel AD (2010) Music, biological evolution, and the brain. *Emerging Disciplines*, ed Bailar M (Rice Univ Press, Houston, TX), pp 91–144.
- Reck D (1997) *Music of the Whole Earth* (Da Capo Press, New York).
- Patel AD (2008) *Music, Language, and the Brain* (Oxford Univ Press, New York).
- Gill KZ, Purves D (2009) A biological rationale for musical scales. *PLoS ONE* 4:e8144.
- McDermott JH, Lehr AJ, Oxenham AJ (2010) Individual differences reveal the basis of consonance. *Curr Biol* 20:1035–1041.
- Huron D (2006) *Sweet Anticipation: Music and the Psychology of Expectation* (MIT Press, Cambridge, MA).
- Huron D (1996) The melodic arch in Western folksongs. *Computing Musical* 10:3–23.
- von Hippel P (2000) Redefining pitch proximity: Tessitura and mobility as constraints on melodic intervals. *Music Percept* 17:315–327.
- Chomsky N (1980) *Rules and Representations* (Columbia Univ Press, New York).
- Podos J, Nowicki S (2004) Performance limits on birdsong. *Nature's Music: The Science of Birdsong*, eds Marler P, Slabberkoorn H (Elsevier, Amsterdam), pp 318–342.
- Schmidt-Nielsen K (1971) How birds breathe. *Sci Am* 225:72–79.
- Riede T, Goller F (2010) Peripheral mechanisms for vocal production in birds—differences and similarities to human speech and singing. *Brain Lang* 115:69–80.
- Nowicki S (1987) Vocal tract resonances in oscine bird sound production: Evidence from birdsongs in a helium atmosphere. *Nature* 325:53–55.
- Nowicki S, Marler P (1988) How do birds sing? *Music Percept* 5:391–426.
- Suthers RA, Goller F, Pytte C (1999) The neuromuscular control of birdsong. *Philos Trans R Soc Lond B Biol Sci* 354:927–939.
- Suthers RA (2004) How birds sing and why it matters. *Nature's Music: The Science of Birdsong*, eds Marler P, Slabberkoorn H (Elsevier, Amsterdam), pp 272–295.
- Sundberg J (1977) The acoustics of the singing voice. *Sci Am* 236:82–84, 86, 88–91.
- Titze IR (2008) The human instrument. *Sci Am* 298:94–101.
- Selfridge-Field E (1995) *Essen Musical Data Package*. Center for Computer Assisted Research in the Humanities, Stanford University, Technical Report No. 1. (CCARH, Menlo Park, CA).
- Kroodma DE, Konishi M (1991) A subsong bird (eastern phoebe, *Sayornis phoebe*) develops normal song without auditory feedback. *Anim Behav* 42:477–487.
- Palmer C (1997) Music performance. *Annu Rev Psychol* 48:115–138.
- Rakowski A (1990) Intonation variants of musical intervals in isolation and in musical contexts. *Psychol Music* 18:60–72.
- Sundberg J (2000) Emotive transforms. *Phonetica* 57:95–112.
- Lindblom B, Sundberg J (2007) The human voice in speech and singing. *Springer Handbook of Acoustics*, ed Rossing T (Springer, New York), pp 669–712.
- Slifka J (2006) Respiratory system pressures at the start of an utterance. *Dynamics of Speech Production and Perception*, eds Divenyi P, et al. (IOS Press, Amsterdam), pp 45–57.
- Alipour F, Scherer RC (2007) On pressure-frequency relations in the excised larynx. *J Acoust Soc Am* 122:2296–2305.
- Myers S, Hansen B (2007) The origin of vowel length neutralization in final position: Evidence from Finnish speakers. *Nat Lang Linguist Theory* 25:157–193.
- Riede T, Suthers RA, Fletcher NH, Blevins WE (2006) Songbirds tune their vocal tract to the fundamental frequency of their song. *Proc Natl Acad Sci USA* 103:5543–5548.
- Watt HJ (1924) Functions of the size of interval in the songs of Schubert and of the Chippewa [i.e., Ojibway] and Teton Sioux [i.e., Lakota] Indians. *Br J Psychol* 14: 370–386.
- Todd N (1985) A model of expressive timing in tonal music. *Music Percept* 3:3–58.
- Eklund R (2008) Pulmonic ingressive phonation: Diachronic and synchronic characteristics, distribution and function in animal and human sound production and in human speech. *J Int Phonetic Assoc* 38:235–324.
- Gans C (1973) Sound production in the Saliencia: Mechanisms and evolution of the emitter. *Am Zool* 13:1179–1194.
- Hartshorne C (1958) The relation of birdsong to music. *Ibis* 100:421–445.
- Baptista LF, Keister RA (2005) Why birdsong is sometimes like music. *Perspect Biol Med* 48:426–443.
- Rothenberg D (2005) *Why Birds Sing: A Journey Into the Mystery of Bird Song* (Basic Books, New York).
- Taylor H (2008) Decoding the song of the pied butcherbird: An initial survey. *Trans-cultural Music Rev* 12:1–30.
- Marler P, Slabberkoorn H (2004) *Nature's Music: The Science of Birdsong* (Elsevier, Amsterdam).
- Kroodma D (2004) *The Singing Life of Birds* (Houghton Mifflin, New York).
- Elliot L (1999) *Music of the Birds: A Celebration of Birdsong* (NatureSound Studio, New York).
- Howard R, Moore A (1984) *A Complete Checklist of Birds of the World* (Macmillan, London).

Supporting information

Tierney et al. 10.1073/pnas.1103882108

SI Experimental Methods

Tracing of Bird Note Frequency Contours on Spectrograms. Birdsongs in our corpus varied in sample rate, depending on their source (sample rates were 48,000, 44,100, 32,000, 22,050, 16,000, 11,025, and 8,000 Hz). Narrow-band spectrograms were made for all birdsongs in SIGNAL version 3.1 (Engineering Design), using a Fast Fourier Transform (FFT) length of 512 points for songs with 48.1 or 44.1 KHz sample rate, 256 points for 32 or 22.05 KHz sample rate, 128 points for 16 or 11.025 KHz sample rate, and 64 points for 8 KHz sample rate. [Thus, the corresponding frequency (and time) resolutions of spectrograms for these seven sample rates were: 94 Hz (11 ms) for 48 KHz sample rate, 86 Hz (12 ms) for 44.1 KHz sample rate, 125 Hz (8 ms) for 32 KHz sample rate, 86 Hz (12 ms) for 22.05 KHz sample rate, 125 Hz (8 ms) for 16 KHz sample rate, 86 Hz (12 ms) for 11.025 KHz sample rate, and 125 Hz (8 ms) for 8 KHz sample rate.] In all spectrograms FFTs were computed every 3 ms, using Hanning windows.

Boundaries of individual birdsong notes were identified using silent intervals and/or sudden pitch jumps between notes on spectrograms. When such boundaries were ambiguous a boundary was set if two further criteria were met: (i) the amplitude waveform approached the noise floor between the candidate notes, and (ii) an audible gap was heard between the notes when the song was played back at half or quarter speed. Using these boundaries, tracing of the fundamental frequency of each note was done automatically using the “Peak” command in SIGNAL (version 3.1), which identifies the frequency with maximum power in successive time columns of a spectrogram (amplitude threshold for peak tracking was set to 40 dB below maximum spectrogram amplitude).

Computation of Average Pitch–Time Contours for Human Song Phrases and Birdsong Notes. For each human song phrase, pitch values were converted to distances from the mean pitch of the phrase (in semitones). Any musical rests (silent gaps) in the phrase were removed, and then each pitch contour was sampled at 50 equally spaced time points. Thus, each human song phrase was converted to a frequency-normalized and time-normalized pitch contour. All contours in the human song corpus were then averaged together, yielding an average human song phrase contour, plotted in Fig. S1A. (Note that for this analysis, all phrases in the human song corpus were used, not just those with five or more notes.)

For birdsong notes, frequency contours were converted to pitch contours [i.e., all frequency points in the contour were converted to semitones from the mean frequency of the note using the formula $ST = 12 \cdot \log_2(F/\text{mean}(F))$, where F is the frequency of a data point in Hz and $\text{mean}(F)$ is the mean frequency of the note]. Construction of an average pitch contour then proceeded in a manner identical to that for human song phrases (i.e., each pitch contour was sampled at 50 equally spaced time points, and all contours in the birdsong corpus were then averaged together, yielding an average birdsong note contour, plotted in Fig. S1B).

Interval Compression Ratio (ICR): Original Analysis. For this analysis, each birdsong note was assigned a discrete frequency in Hz, based on the median frequency of the note, and pitch intervals (in semitones) were computed between successive notes in a song, using the equation: interval size = $12 \cdot \log_2(F_2/F_1)$, where F_1 is the median frequency of the first note of the interval, and F_2 is the median frequency of the second note of the interval.

For each human song phrase and birdsong, the ICR was computed as follows. First, notes were randomly reordered in time, and the average absolute pitch interval size of the resulting sequence was computed. This procedure was repeated 100 times, and the mean of these average values was used as the numerator in the ICR value. The denominator was the average absolute interval size when the notes were in their original order. Fig. S2 shows a histogram of the resulting avian and human ICR values. In choosing song phrases for the human ICR analysis, phrases with rests were excluded, as were any human song phrases that could not be matched with a birdsong of corresponding length. Phrases with no pitch variation were also excluded, since our criteria for selecting birdsongs for this study had excluded birdsongs with minimal pitch variation (*Experimental Methods*). The resulting sample size for human song phrases was 34,586.

ICR: Monte Carlo Analysis. To compare the ICRs of human and avian song in a way that matched the sample sizes of ICR values in the two domains, for each of the 80 birdsongs we randomly selected a human song phrase of the same number of notes. (Human song phrases including rests or no pitch variation were excluded, and two birdsongs were long enough that there were no human song phrases of the same length; these birdsongs were excluded from the analysis.) We tabulated these human ICR values and stored the median value of this distribution. We then repeated this procedure 1,000 times and created a distribution of median human ICR values. This allowed us to compute a P value for observing a median human ICR value equal to or smaller than the median avian ICR value of 1.21. According to these 1,000 draws, the smallest median human ICR value was 1.28, meaning that $P = 0$. Hence humans do indeed seem to be more biased toward small pitch intervals in their song melodies than are birds, as predicted by the motor constraint hypothesis.

SI Results

ICR Analysis of Chinese Songs. The significantly greater human vs. avian ICR values in the original and Monte Carlo analyses were replicated for the Chinese songs analyzed separately, when Chinese songs with at least six scale tones were analyzed ($N = 3,999$ phrases, representing 47% of all phrases eligible for ICR analysis in the Chinese corpus, i.e., phrases without rests and with pitch variation). Pentatonic songs (i.e., songs with five scale tones) were very common in the Chinese corpus, and had ICR values that were not significantly different from avian ICR values. This is likely because songs built from pentatonic scales tend to have large pitch jumps between their notes simply by virtue of their scale structure (1).

1. Han Se, Sundararajan J, Bowling DL, Lake J, Purves D (2011) Co-variation of tonality in the music and speech of different cultures. *PLoS One* 6(5):e20160.

Table S1. Regions, number of songs, and number of melodic phrases in the Essen folk song database

Region	No. of songs	No. of phrases
Germany	5,252	30,977
China	2,041	10,258
Luxembourg	621	3,563
France	589	3,113
Nova Scotia	152	885
Austria	130	817
Yugoslavia	119	303
Switzerland	99	550
Ireland	63	343
Netherlands	62	313
Russia	51	213
Scotland	51	422
Hungary	46	189
Czech Republic	43	165
Romania	28	114
Belgium	26	130
Poland	25	117
Sweden	12	52
Mexico	11	77
United States	10	61
Denmark	9	41
England	4	32
Italy	4	18
Norway	2	9
Iceland	2	10
Greece	1	3
Finland	1	4
Brazil	1	2
Canada	1	4
Japan	1	7
India	1	6
Indonesia	1	14
Miscellaneous regions (unclassified)	8	87
Total	9,467	52,899

Table S2. Families, common names, and scientific names of bird species studied

Family	Common name	Species name
Acanthizidae	Striated fieldwren	<i>Calamanthus fuliginosus</i>
Aegithinidae	Common lora	<i>Aegithina tiphia</i>
Alaudidae	Crested lark	<i>Galerida cristata</i>
Alaudidae	Woodlark	<i>Lullula arborea</i>
Atrichornithidae	Noisy scrub-bird	<i>Atrichornis clamosus</i>
Bombycillidae	Cedar waxwing	<i>Bombycilla cedrorum</i>
Campephagidae	Madagascar cuckoo-shrike	<i>Coracina cinerea</i>
Campephagidae	Scarlet minivet	<i>Pericrocotus flammeus</i>
Cardinalidae	Northern cardinal	<i>Cardinalis cardinalis</i>
Cardinalidae	Rose-breasted grosbeak	<i>Pheucticus ludovicianus</i>
Certhiidae	Eurasian treecreeper	<i>Certhia familiaris</i>
Certhiidae	Rusty-flanked treecreeper	<i>Certhia nipalensis</i>
Chloropseidae	Lesser green leafbird	<i>Chloropsis cyanopogon</i>
Cinclidae	American dipper	<i>Cinclus mexicanus</i>
Cisticolidae	Gray-capped warbler	<i>Eminia lepida</i>
Cisticolidae	Rattling cisticola	<i>Cisticola chiniana</i>
Colluricinclidae	Gray shrike-thrush	<i>Colluricincla harmonica</i>
Cracticidae	Pied butcherbird	<i>Cracticus nigrogularis</i>
Dicaeidae	Fire-breasted flowerpecker	<i>Dicaeum ignipectus</i>
Dicruridae	Lesser racket-tailed drongo	<i>Dicrurus remifer</i>
Dicruridae	Pale blue monarch	<i>Hypothymis puella</i>
Emberizidae	Field sparrow	<i>Spizella pusilla</i>
Emberizidae	White-throated sparrow	<i>Zonotrichia albicollis</i>
Estrildidae	Common waxbill	<i>Estrilda astrild</i>
Eupetidae	Chiming wedgebill	<i>Psophodes occidentalis</i>
Fringillidae	European greenfinch	<i>Carduelis chloris</i>
Fringillidae	Oriental greenfinch	<i>Carduelis sinica</i>
Hirundinidae	Northern rough-winged swallow	<i>Stelgidopteryx serripennis</i>
Icteridae	Baltimore oriole	<i>Icterus galbula</i>
Icteridae	Eastern meadowlark	<i>Sturnella magna</i>
Malaconotidae	Black-crowned tchagra	<i>Tchagra senegala</i>
Malaconotidae	Brubru shrike	<i>Nilaus afer</i>
Maluridae	Splendid fairywren	<i>Malurus splendens</i>
Maluridae	Superb fairywren	<i>Malurus cyaneus</i>
Meliphagidae	Giant honeyeater	<i>Gymnomyza viridis</i>
Meliphagidae	Kauai O'o (extinct)	<i>Moho braccatus</i>
Mimidae	Northern mockingbird	<i>Mimus polyglottos</i>
Monarchidae	Black-faced monarch	<i>Monarcha melanopsis</i>
Motacillidae	Tree pipit	<i>Anthus trivialis</i>
Muscicapidae	Nightingale	<i>Luscinia megarhynchos</i>
Muscicapidae	White-browed bush robin	<i>Tarsiger indicus</i>
Nectariniidae	Bronze sunbird	<i>Nectarinia kilimensis</i>
Nectariniidae	Crimson sunbird	<i>Aethopyga siparaja</i>
Oriolidae	Indian golden oriole	<i>Oriolus kundoo</i>
Pachycephalidae	Golden whistler	<i>Pachycephala pectoralis</i>
Paridae	Black-capped chickadee	<i>Poecile atricapillus</i>
Paridae	Turkestan tit	<i>Parus bokharensis</i>
Parulidae	Chestnut-sided warbler	<i>Dendroica pensylvanica</i>
Parulidae	Yellow warbler	<i>Dendroica petechia</i>
Passeridae	Mauritius fody	<i>Foudia rubra</i>
Petroicidae	New Zealand robin	<i>Petroica australis</i>
Petroicidae	South Island tomtit	<i>Petroica macrocephala</i>
Peucedramidae	Olive warbler	<i>Peucedramus taeniatus</i>
Ploceidae	Spectacled weaver	<i>Ploceus ocularis</i>
Poliptilidae	Long-billed gnatwren	<i>Ramphocaenus melanurus</i>
Prunellidae	Brown accentor	<i>Prunella fulvescens</i>
Pycnonotidae	Orange-spotted bulbul	<i>Pycnonotus bimaculatus</i>
Pycnonotidae	Spectacled bulbul	<i>Pycnonotus erythrophthalmos</i>
Regulidae	Firecrest	<i>Regulus ignicapilla</i>
Regulidae	Golden-crowned kinglet	<i>Regulus satrapa</i>
Remizidae	Eurasian penduline-tit	<i>Remiz pendulinus</i>
Rhipiduridae	Pied fantail	<i>Rhipidura javanica</i>
Rhipiduridae	White-browed fantail	<i>Rhipidura aureola</i>

Table S2. Cont.

Family	Common name	Species name
Sittidae	Rock nuthatch	<i>Sitta neumayer</i>
Sturnidae	Common myna	<i>Acridotheres tristis</i>
Sylviidae	Japanese bush-warbler	<i>Cettia diphone</i>
Sylviidae	Willow warbler	<i>Phylloscopus trochilus</i>
Thraupidae	Summer tanager	<i>Piranga rubra</i>
Thraupidae	Variable seedeater	<i>Sporophila corvina</i>
Timaliidae	Hwamei	<i>Garrulax canorus</i>
Timaliidae	Spot-breasted laughingthrush	<i>Garrulax merulinus</i>
Troglodytidae	Carolina wren	<i>Thryothorus ludovicianus</i>
Troglodytidae	Canyon wren	<i>Catherpes mexicanus</i>
Turdidae	Slaty-backed nightingale thrush	<i>Catharus fuscaer</i>
Turdidae	White-browed shortwing	<i>Brachypteryx montana</i>
Vangidae	Rufous vanga	<i>Schetba rufa</i>
Viduidae	Pin-tailed whydah	<i>Vidua macroura</i>
Vireonidae	Red-eyed vireo	<i>Vireo olivaceus</i>
Vireonidae	Green shrike-vireo	<i>Vireolanius pulchellus</i>
Zosteropidae	Mascarene white-eye	<i>Zosterops borbonicus</i>

Audio File S1. Human folk song (Fig. 1 A and B, main text).

[Audio File S1](#)

Audio File S2. Field sparrow song (Fig. 2A, main text).

[Audio File S2](#)

Audio File S3. Eurasian treecreeper song (Fig. 2C, main text). In this example background birds are audible, and their sounds are visible in the spectrogram as a band of energy around 4 KHz. When listening to the song, note that the first note of the treecreeper song (which starts at ≈ 300 ms and ≈ 6.7 KHz) is immediately preceded by a separate background bird note (at ≈ 250 ms and ≈ 3.5 KHz).

[Audio File S3](#)

Audio File S4. Eurasian treecreeper song, played at quarter speed. Note that the rising and falling notes between $\approx 1,350$ and $1,450$ ms on this spectrogram (i.e., notes 6 and 7 out of 17 notes total) sound like one note when played at normal speed but can be heard as two distinct notes at quarter speed. Additionally, in the acoustic waveform, the two notes are separated by a brief period in which waveform amplitude drops to the level of the ambient noise. This illustrates how slowing down the songs, in combination with inspection of the waveform and spectrogram, was used to decide on boundaries between notes.

[Audio File S4](#)

Audio File S5. Summer tanager song (Fig. 2D, main text).

[Audio File S5](#)

Audio File S6. Summer tanager song, played at half speed.

[Audio File S6](#)