

# Overtone-based pitch selection in hermit thrush song: Unexpected convergence with scale construction in human music

Emily L. Doolittle<sup>a,1</sup>, Bruno Gingras<sup>b,1</sup>, Dominik M. Endres<sup>c,d</sup>, and W. Tecumseh Fitch<sup>b,2</sup>

<sup>a</sup>Department of Music, Cornish College of the Arts, Seattle, WA 98121; <sup>b</sup>Department of Cognitive Biology, Faculty of Life Sciences, University of Vienna, Vienna A-1090, Austria; <sup>c</sup>Theoretical Neuroscience Group, Philipps University of Marburg, 35032 Marburg, Germany; and <sup>d</sup>Section for Computational Sensomotorics, Hertie Institute for Clinical Brain Research, Center for Integrative Neuroscience, Bernstein Center for Computational Neuroscience, and University Clinic Tübingen, 72076 Tübingen, Germany

Edited by Dale Purves, Duke University, Durham, NC, and approved October 8, 2014 (received for review April 3, 2014)

Many human musical scales, including the diatonic major scale prevalent in Western music, are built partially or entirely from intervals (ratios between adjacent frequencies) corresponding to small-integer proportions drawn from the harmonic series. Scientists have long debated the extent to which principles of scale generation in human music are biologically or culturally determined. Data from animal “song” may provide new insights into this discussion. Here, by examining pitch relationships using both a simple linear regression model and a Bayesian generative model, we show that most songs of the hermit thrush (*Catharus guttatus*) favor simple frequency ratios derived from the harmonic (or overtone) series. Furthermore, we show that this frequency selection results not from physical constraints governing peripheral production mechanisms but from active selection at a central level. These data provide the most rigorous empirical evidence to date of a bird song that makes use of the same mathematical principles that underlie Western and many non-Western musical scales, demonstrating surprising convergence between human and animal “song cultures.” Although there is no evidence that the songs of most bird species follow the overtone series, our findings add to a small but growing body of research showing that a preference for small-integer frequency ratios is not unique to humans. These findings thus have important implications for current debates about the origins of human musical systems and may call for a reevaluation of existing theories of musical consonance based on specific human vocal characteristics.

music | birdsong | overtones

Many human musical scales, including the diatonic major scale prevalent in Western music, are built partially or entirely from intervals (ratios between adjacent frequencies) corresponding to small-integer ratios drawn from the harmonic series (1). A long-running debate concerns the extent to which principles underlying the structure of human musical scales derive from biological aspects of auditory perception and/or vocal production or are historical cultural “accidents” (2–4). The songs of nonhuman animals, such as birds or whales, potentially offer a valuable perspective on this debate. On the one hand, features of human music that are culturally bound, or dependent on specific characteristics of the human voice or auditory system, should be absent in animal vocalizations. On the other hand, aspects of human music observed in the vocalizations of other species seem likely to be partially determined by general physical or biological constraints rather than solely by cultural practices. Such shared features would complement recent research suggesting that common motor constraints shape both human song and that of some bird species (5).

The physical principles underlying vocal production in songbirds are well understood (6–10) and do not differ fundamentally from those of other vertebrates. Sound is produced by tissue vibrations in the syrinx, a bird-specific organ located at the base of the trachea. Flow-driven vibrations of fleshy membranes within

the syrinx (in songbirds, the medial and lateral labia) generate a periodic source signal that is filtered by the air column within the trachea and mouth and then emitted to the environment. These principles are important in formulating various alternative hypotheses considered below.

Naturalists have long wondered whether birdsong could be said to have musical properties (11–13). However, early studies on pitch selection tended to be anecdotal, based on a small sample size, or lacking in analytical rigor. Two more recent studies specifically comparing pitch selection in bird song and human musical scales concluded that birdsong does not make preferential use of musical intervals found in commonly used Western musical scales (14, 15). However, because these studies each only examined one species [the white-throated sparrow (*Zonotrichia albicollis*) and the nightingale wren (*Microcerculus philomela*), respectively], a conclusion that birdsong in general does not exhibit musical properties seems premature. Indeed, other studies have shown preferential use of consonant intervals in tropical boubou shrikes (*Laniarius aethiopicus*) (16) and musician wrens (*Cyphorhinus arada*) (17), although in the first case no rigorous statistical analysis was presented.

Here, we investigated songs of the hermit thrush (*Catharus guttatus*), a medium-sized North American songbird whose famously “musical”-sounding song has attracted the attention of ornithologists and musicians alike (18) but has not yet been subjected to detailed pitch analysis. Its songs are composed of

## Significance

The song of the hermit thrush, a common North American songbird, is renowned for its apparent musicality and has attracted the attention of musicians and ornithologists for more than a century. Here we show that hermit thrush songs, like much human music, use pitches that are mathematically related by simple integer ratios and follow the harmonic series. Our findings add to a small but growing body of research showing that a preference for small-integer ratio intervals is not unique to humans and are thus particularly relevant to the ongoing nature/nurture debate about whether musical predispositions such as the preference for consonant intervals are biologically or culturally driven.

Author contributions: E.L.D. and W.T.F. designed research; E.L.D. and W.T.F. performed research; E.L.D., B.G., and D.M.E. analyzed data; and E.L.D., B.G., D.M.E., and W.T.F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

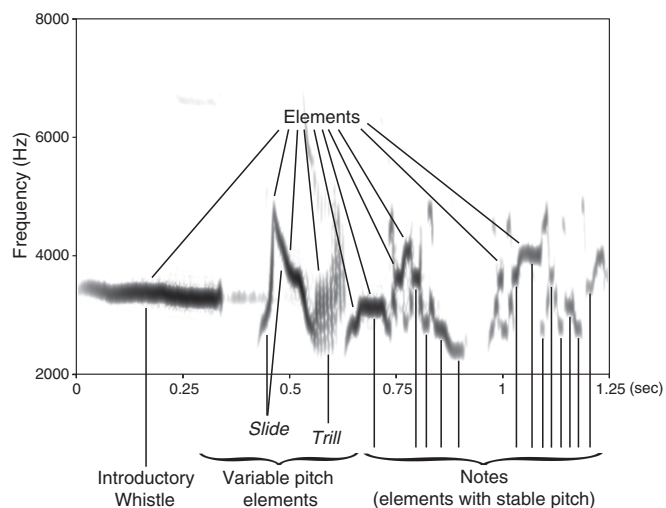
<sup>1</sup>E.L.D. and B.G. contributed equally to this work.

<sup>2</sup>To whom correspondence should be addressed. Email: tecumseh.fitch@univie.ac.at.

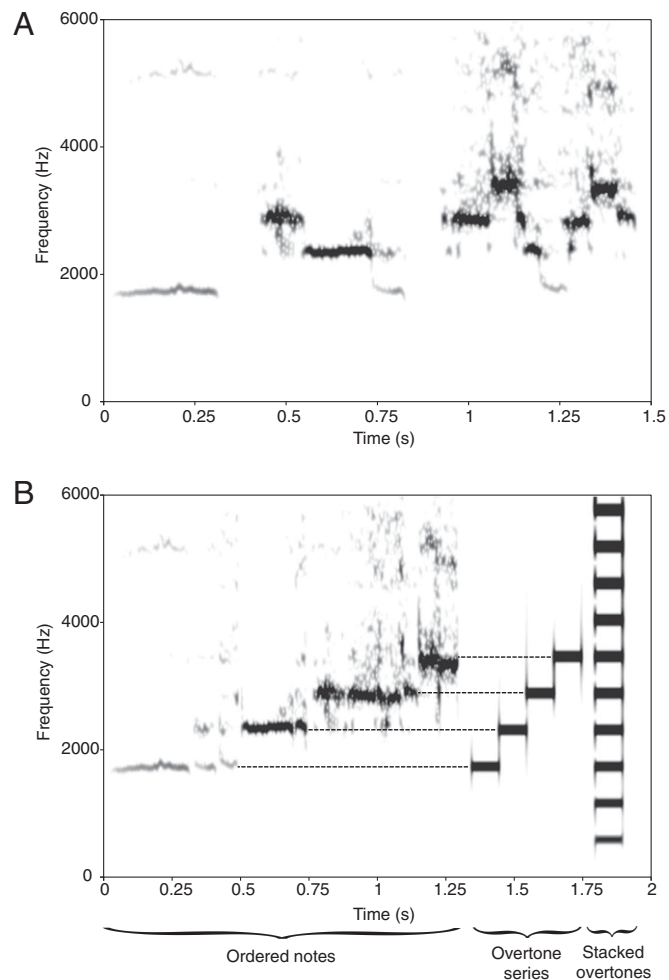
This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1406023111/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1406023111/-DCSupplemental).

elements (the smallest unit of song construction, seen as continuous uninterrupted traces on spectrograms) that may exhibit either a variable pitch, such as trills and slides, or a stable pitch—pure, non-frequency-modulated, “flutelike” sounds. These stable sounds, which we refer to as “notes” (Fig. 1), are characterized by strong fundamental frequencies and very weak higher harmonics, making them ideally suited for an analysis of pitch relationships (15). Males typically sing 6–10 different song types, defined as nearly identical sequences of elements, durations, and frequencies. In a number of early- and mid-20th-century studies, hermit thrush song was variously attributed with use of major, minor, and pentatonic scales (19, 20) and claimed to follow the overtone series (21). However, these early studies again suffered from small sample sizes and anecdotal reporting and were not based on rigorous acoustic analysis. More recent hermit thrush studies have focused on regional differences and song-type ordering, rather than pitch selection (22, 23).

Here we tested the overtone hypothesis, which predicts that the frequencies of the individual song notes are integer multiples (harmonics) of an implied (but not actually sung) base frequency (hereafter  $f_i$ ). This hypothesis seems plausible because, unlike some previous claims, it does not attribute human-specific music-theoretical concepts to hermit thrush song. Moreover, the subjective impression of trained musicians listening to hermit thrush songs (played at one-sixth of the original speed to shift the speed and frequency of the songs into a range more suitable for human hearing) was that most notes indeed seemed to follow an overtone series (see Fig. 2 and [Audio File S1](#) for the corresponding sound example). However, determining whether a set of notes are harmonics of a frequency not present in the set requires a rigorous procedure to estimate and evaluate  $f_i$ . To this end, we used two different statistical approaches, an ordinary least-squares regression model and a generative Bayesian estimator. Both approaches were used to test the hypothesis that a song is an exchangeable sequence of frequencies that are integer multiples of some implied  $f_i$ , versus the null hypothesis that songs are generated by drawing frequencies out of a random log-normal distribution (see *Materials and Methods* for details). By using a Bayesian approach in addition to the least-squares regression model we evaluate whether our analyses represent a rigorous test of our overtone hypothesis and not simply a post hoc explanation



**Fig. 1.** Song of the hermit thrush (*C. guttatus*). One song type of a single male hermit thrush, illustrating the various elements that can be observed in songs of this species. Only “notes” (elements with stable pitch) were analyzed in this study because the other element types have no clearly defined or measurable pitch.



**Fig. 2.** Frequency distribution of a hermit thrush song compared with an overtone series. (A) Notes of a hermit thrush song. (B) The same notes rearranged in ascending order to show how they correspond to overtones 3, 4, 5, and 6 of an overtone series fitted to the frequencies corresponding to these notes (the complete stacked overtone series is shown on the right).

that minimizes an error measure by “memorizing” the data. These properties make the Bayesian evaluation statistically more rigorous than least-squares fitting.

## Results and Discussion

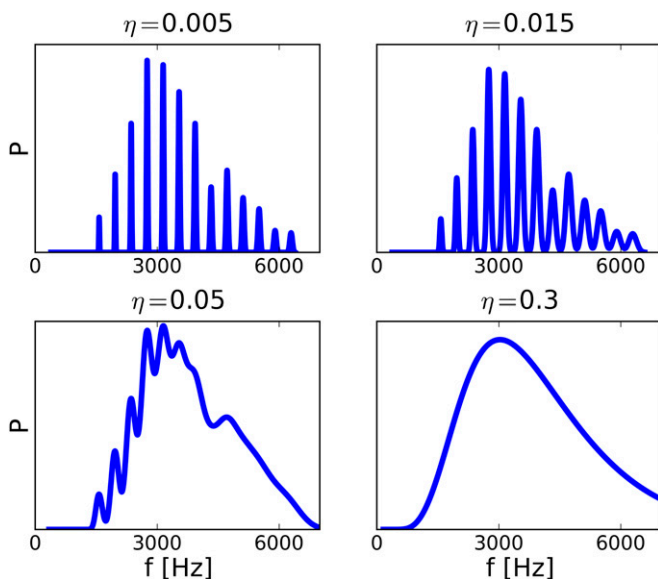
From our collection of 114 song types produced by 14 male hermit thrushes across a wide time range, and spanning North America, we analyzed all song types containing 10 or more notes for which a single stable frequency could accurately be determined (shown as “elements with stable pitch” in Fig. 1), for a total of 71 song types. The prominent first element of each song type, the “introductory whistle,” was omitted because its pitch varies over time, often rising or falling by up to 200 cents or more, and thus cannot accurately be assigned to a single frequency.

Using the least-squares regression model we found that the frequencies of the notes from 57 of 71 songs followed a distribution that is significantly closer to an overtone series than to a random log-normal distribution, indicating that notes from these songs typically approximated integer multiples of an inferred  $f_i$ . According to the Bayesian estimator, 61 of 71 songs were significantly overtone-related. Although these two statistical approaches are conceptually quite different, both agreed in their classification of 61 of the 71 songs, 54 of which were classified as harmonic by both models, and 7 as nonharmonic (the Bayesian

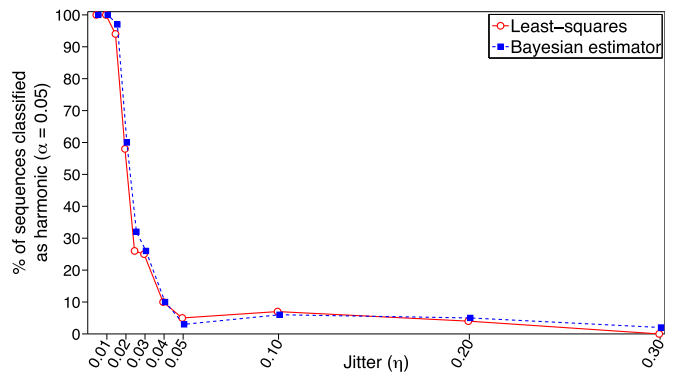
estimator classified an additional 7 songs as harmonic, and the least-squares regression model an additional 3). Moreover, both approaches agreed on the  $f_i$  values for 57 songs (sound examples of songs classified as harmonic and nonharmonic can be found in [Audio Files S2–S13](#) and [Table S1](#)).

To ascertain that our statistical models correctly classify as harmonic those songs whose note frequencies closely approximated an overtone series while rejecting nonharmonic songs, we evaluated the validity of the Bayesian estimator and least-squares regression model on a computer-generated “ground truth” dataset, consisting of 1,100 sequences comprising 14 notes each (the average number of notes in hermit thrush songs from our database). Note frequencies in these sequences were drawn from distributions spanning the continuum from strictly harmonic to a distribution with a frequency “jitter” or logarithmic SD ( $\eta$ ) from an exact overtone series equal to 0.3 (visually undistinguishable from a log-normal distribution; see Fig. 3 and *Materials and Methods* for details). Both regression and Bayesian approaches classified 10% or fewer of all sequences generated from a distribution with  $\eta \geq 0.04$  as harmonic (at a significance threshold  $\alpha = 0.05$ ), whereas more than 90% of all sequences generated from a distribution with  $\eta \leq 0.015$  were classified as harmonic (Fig. 4). Note that a distribution with  $\eta = 0.015$  still exhibits a clear harmonic structure (Fig. 3). In comparison, human singers singing the well-known song “Happy Birthday” show an empirical deviation of  $\eta = 0.014$  from the intended frequency (*Materials and Methods*).

To further validate our statistical models we analyzed 10 commercial recordings of the alphorn, an instrument made out of a holeless, conical bore tube that is physically capable only of producing notes from the overtone series above a base frequency produced by the entire length of the instrument. Both statistical models identified all alphorn recordings as overtone-related and correctly identified  $f_i$  for each example. As expected, the alphorn recordings yielded a better least-squares fit to the overtone series than the hermit thrush songs (Mann–Whitney  $U = 80$ ,  $P < 0.001$ ), because the alphorn deviates only minimally from a pure overtone series. However, the residual error of the least-squares



**Fig. 3.** Probability densities of frequency distributions with various degrees of jitter. Example of frequency distribution densities from which the ground truth dataset was generated. For small values of jitter ( $\eta$ ), the densities exhibit regularly spaced, distinct peaks. Sequences of notes drawn from these densities thus display an overtone structure. For larger values of  $\eta$ , the peaks begin to overlap, and the overtone structure disappears.



**Fig. 4.** Sensitivity analysis. Percentage of 14-note sequences classified as overtone-related (at a significance threshold  $\alpha = 0.05$ ) for the least-squares regression model (continuous line and open circles) and for the Bayesian estimator (dotted line and closed squares) for jitter values ( $\eta$ ) ranging from 0.005 to 0.3.

fit for 57 of 71 hermit thrush songs was less than twice the mean residual error of the fit for the alphorn recordings, and 52 of these 57 songs were classified as significantly harmonic by the least-squares regression model. This indicates that most hermit thrush songs from our sample were only moderately more “out-of-tune” than alphorn recordings.

One possible explanation for the strong bias toward overtone-related pitches in hermit thrush songs could be that the birds couple their vocal fundamental frequency to the resonances of their vocal tract, in a manner similar to wind instruments like the alphorn. If this is true, birds would generate a harmonic series using the acoustic resonances of their tube-like trachea (24), and the observed frequency distribution would follow from basic physical production constraints. However, there is considerable evidence against the source–tract coupling in birds (25, 26) that would be required by this hypothesis. Equally critically, the implied fundamentals for the hermit thrush songs were drastically lower than those predicted using measured tracheal lengths. We calculated the predicted  $f_i$  for a half-open air-filled tube using the formula  $f_i = c/4 \times L$ , where  $c$  is the speed of sound in warm, moist air (350 m/s) and  $L$  is the tube length (27). Fully stretched, the lengths of two hermit thrush tracheas were 32 and 35 mm, predicting  $f_i$  values of 2,734 and 2,500 Hz (27), far higher than our estimated values, which all fell between 180 and 720 Hz. Furthermore, hermit thrush songs often contain occasional non-overtone notes, as well as smooth frequency glides over a large frequency range, which would be impossible with a fixed-length trachea if source–tract coupling were present. All of these data are inconsistent with the physical constraints on peripheral production mechanisms required by the “alphorn” hypothesis.

Alternatively, a bird could theoretically select specific harmonics from a fixed-pitched source by varying its vocal tract filter, as occurs during overtone or “throat” singing in humans (28, 29). However, this would require both a very low source fundamental frequency and a vocal tract flexible enough to allow the bird to pick out many different harmonics of that source, neither of which conditions can plausibly be met by hermit thrush vocal anatomy.

Thus, the production of notes from the harmonic series does not result directly from the physics of the vocal apparatus but instead seems to derive from voluntary central control of muscular and neural parameters. Supporting the idea that it is possible for songbirds to learn to select pitches from the overtone series, white-crowned sparrows, whose songs do not normally follow the overtone series, can learn to sing hermit thrush songs in experimental conditions (30).

Our results show that pitches within *C. guttatus* songs are related by small-integer ratio intervals from the overtone series,

indicating that pitch selection in these songs obeys mathematical constraints found within common human musical scales. Because these constraints are found in hermit thrush songs from across North America, over a span of more than 50 y, our data suggest that this is a species-typical characteristic of hermit thrush song. Although the presence of small-integer ratio intervals is neither a necessary nor a sufficient condition for something to be considered music [e.g., some music does not contain any defined pitch intervals, whereas a doorbell or other nonmusical signal may ring a perfect fifth (3:2) or other simple integer ratio], these intervals have a preferential status in practically all known human musical cultures (31) and thus may be considered a characteristic feature of music.

Although pitch selection in birdsong has only been studied quantitatively in a few species, other bird species, such as Java sparrows (*Padda oryzivora*) (32), European starlings (*Sturnus vulgaris*) (33), and pigeons (*Columba livia*) (34) have been shown to discriminate between consonance and dissonance, and newly hatched domestic chicks (*Gallus gallus*) display a preference for consonant intervals (35). Outside the avian kingdom, octave generalization has been shown in rhesus macaques (*Macaca mulatta*) (36), and pairs of dengue vector mosquitoes (*Aedes aegypti*) converge on buzzing frequencies that are a perfect fifth apart before mating (37). Given that few rigorous studies have concentrated on pitch selection or perception in nonhuman animals, our findings lead us to predict that future studies may show a preference for consonant intervals in more species.

A leading hypothesis for human consonance preference suggests that early (even prenatal) exposure to the harmonic-rich human voice, combined with the specific characteristics of the human vocal tract, provides an acquired “template” for musical attractiveness (4, 38, 39). Because hermit thrushes’ notes, like those of most birdsong, lack strong harmonics (Fig. 1), this explanation cannot apply in this species. Thus, early exposure to broad-band harmonic sounds (like the human voice) is not necessary for an organism to favor notes chosen from the harmonic series. Why, then, would hermit thrushes consistently choose to sing using pitches from the harmonic series? One possibility is that the acoustical predictability of overtone series provides an objective yardstick for females evaluating a singing male’s pitch accuracy. Another possibility, not mutually exclusive, is that frequencies related by small-integer ratios may be more easily remembered or processed in the auditory system, representing a form of sensory bias that may characterize both humans (31) and at least some species of birds.

Our results, combined with the above-mentioned evidence for preferential use of consonant intervals in other animal species,

support the assertion that some aspects of human scales may be partially based on shared biological principles. More generally, these results, along with recent work on rhythmic entrainment in animals (40, 41), suggest that a number of perceptual and motor mechanisms providing the biological bases for human music may be shared with some other species and call for a reevaluation of long-held assumptions about the species-specific nature and origin of human musical preferences.

## Materials and Methods

**Recordings and Acoustic Analysis.** High-quality recordings of hermit thrush songs from 14 birds were acquired from three sources: the Borror Laboratory of Bioacoustics, Kevin J. Colver (Kevin J. Colver Productions, Elk Ridge, UT), and Bernie Krause (Wild Sanctuary, Glen Ellen, CA). Recordings and passages in which a poor signal-to-noise ratio or overlapping vocalizations prevented accurate analysis of the songs, or in which the focal bird sang fewer than 12 songs, were discarded. The longest available uninterrupted passage of the hermit thrush’s vocalizations was selected from each recording. In the case of one bird there were two suitable uninterrupted passages and we used data from both. Acoustic analysis was performed using the software Praat (42) and custom Praat macros (written by W.T.F.). Recordings for each bird were segmented into individual songs, defined as bouts of vocalization separated by more than 2 s of silence. These extracted songs were saved as separate files, numbered consecutively, and labeled according to start time and duration. The accuracy of automated extraction of the fundamental pitch of each note was verified by ear and by visual inspection of spectrograms with overlaid pitch tracks.

**Song Types and Ordering in Song Series.** Hermit thrush songs are made up of small number of song types (22, 23); to avoid pseudoreplication we analyzed only the first appearance of each song type of each bird. Song types were defined as identical sequences of elements, as determined by visual inspection of the sonograms, and by comparing the average frequencies of the introductory whistles and the frequencies of the postintroductory whistle notes of each song. Each of our birds sang between 6 and 10 song types, for a total of 114 different song types (Table 1). None of these song types was shared between birds. Each bird typically cycled through all his song types within about 14 songs. Song types were never immediately repeated. The largest observed repeat interval was 32 songs and the shortest 2 songs. Individual birds varied in the predictability with which they presented their song types.

**Pitch Extraction.** Praat’s autocorrelation-based pitch extraction algorithm was used to estimate the average fundamental frequency (“F0” or “pitch” hereafter) of each note with a steady, determinable pitch and that lasted for at least 10 ms. The minimum F0 of a note measured from any bird was 1,545 Hz and the maximum was 7,925 Hz. The mean range for each song type was 1,190 cents (very close to an octave, which corresponds to 1,200 cents). The mean range covered by each bird (including all song types) was 2,537 cents, representing slightly more than two octaves, with a minimum range of 2,201 cents and a maximum of 3,002 cents.

**Table 1. Source, recording date, location, duration, and number of songs for each sample**

Sample	Source	Date	Location	Length, s	No. of songs	No. of song types
1	Krause	Unknown	Southeast Alaska	120	30	6
2	Borror	07/11/1970	Upper Peninsula, MI	72	12	6
3	Borror	07/19/1955	Muscongus, ME	107	18	9
4	Borror	05/26/1988	Upper Peninsula, MI	249	65	9
5	Borror	07/15/1956	Muscongus, ME	98	23	9
6	Borror	07/19/1956	Muscongus, ME	119	22	10
7	Borror	06/10/1997	Homer, AK	332	66	7
8	Borror	07/01/1958	Nobleboro, ME	108	18	8
9	Borror	07/28/1953	Millinocket, ME	95	18	6
10	Borror	07/07/1965	Lewisburg, OR	104	26	7
11	Borror	07/09/1967	North Rim, AZ	114	18	9
12	Colver	06/28/1996	Kamas, UT	73	13	10
13	Colver	06/28/1996	Kamas, UT	212	40	9
14	Colver	2007	Homer, AK	171; 258	38; 58	9

Borror, Borror Laboratory of Bioacoustics; Colver, Dr. Kevin J. Colver; Krause, Dr. Bernie Krause.

**Estimating the Harmonicity of the Songs.** To estimate the harmonicity of the hermit thrush songs we used two approaches, one based on an ordinary least-squares regression model and a second on a Bayesian estimator of harmonicity. Both models assume that  $f_i$  lies between 100 and 1,000 Hz (although identical results were obtained when considering a range from 20 to 5,000 Hz) and both consider only the first 16 overtones, given that higher overtones are less clearly separated on a logarithmic basis.

**Ordinary Least-Squares Regression Model.** The ordinary least-squares regression model makes no particular assumptions about the distribution of  $f_i$  besides those outlined above. For each song, a best-fitting base frequency  $f_i$  is computed by finding the value of  $f_i$  that minimizes the mean square error (logarithm of the frequency deviation) of a linear regression that fits the pitches of each note in a song to integer multiples of  $f_i$ . To test the hypothesis H1 that a song is an exchangeable sequence of frequencies that are integer multiples of some implied  $f_i$  versus the null hypothesis H0 that songs are generated by drawing frequencies out of a log-normal distribution, we generated, for each song, 1,000 randomly generated songs with the same number of notes as the actual song, but using frequencies taken from a log-normal distribution (restricted between 0 and 10,000 Hz) modeling the frequency distribution of all hermit thrush songs in our sample. For each of these randomly generated songs, a best fit for  $f_i$  was computed using the least-squares approach described above. If fewer than 5% of the randomly generated songs showed a better fit (smaller mean square error) to an overtone series than the actual song, H0 was rejected at a significance threshold of 0.05.

**Generative Bayesian Model.** The generative Bayesian model computes the posterior distribution of the implied base  $f_i$  in a manner analogous to performing a regression analysis over the measured frequencies of all individual song notes (dependent variable) relative to those predicted from integer multiples (predictor) of all possible  $f_i$  between 100 Hz and 1,000 Hz. Because the predictor values (represented by the integer multiples 1, 2, 3, 4, ..., with  $f_i$  being the frequency associated with the multiple 1) are unknown, this is an instance of a latent variable problem. The model therefore generates distributions over predictor values, rather than point estimates.

A song is parameterized by a (strictly positive) base frequency  $b$  and an overtone distribution  $q = (q_1, \dots, q_m)$ , where  $q_m$  is the probability of observing overtone  $m$  in a song,  $M$  is the maximal possible overtone and  $\sum_{m=1}^M q_m = 1$ . To generate a song, we drew  $K$  notes  $n_k \in \{1, \dots, M\}$  out of  $q$ , where  $k = 1, \dots, K$  is the sequence index. These notes are latent variables, because they are not directly observable. In a perfectly harmonic song, the notes and the corresponding observable pitch frequencies  $f_k$  would be related by

$$f_k = b \cdot n_k. \quad [1]$$

To allow for a certain amount of random frequency deviation (hereafter jitter), which might be attributed to inaccurate singing or measurement errors, we included this jitter  $\rho$  multiplicatively:

$$f_k = b \cdot n_k \cdot \rho, \quad [2]$$

where  $\rho$  is drawn from a log-normal distribution with mean 0 and SD  $\sigma$ . Factorizing the jitter from the "ideal" frequency  $b \cdot n_k$  allows us to use a frequency-independent relative jitter model, because the frequency dependence is already captured by the first two factors on the right-and side of Eq. 2.

To complete the model, we specified priors on  $q$  and  $b$ , because they are a priori unknown and are thus treated as random variables. Because the notes  $n_k$  are multinomial variables (each note can take exactly one of  $M$  values),  $q$  is a multinomial distribution. The canonical prior on a multinomial distribution is the Dirichlet distribution (see ref. 43 for details). We used a symmetric Dirichlet distribution with concentration parameter  $\alpha$ . To choose a prior on  $b$ , note that there is a scaling ambiguity in the model: Substituting  $b$  by  $b/z$  and  $n_k$  by  $z \cdot n_k$  with  $z$  integer and positive would lead to the same observable frequencies  $f_k$ . To remove this ambiguity, we limited the range of  $b$  to  $b \in [b_{\min}, b_{\max}]$  with a density

$$\rho(b) \propto b, \quad [3]$$

that is, we prefer higher base frequencies linearly. The constant of proportionality in Eq. 3 was chosen to ensure normalization:  $\int_{b_{\min}}^{b_{\max}} db \rho(b) = 1$ .

For base frequency inference we computed the posterior distribution of  $b$  given a song using standard Bayesian approaches (43). This computation involves an intractable integral over  $q$ , which we approximate by its value at the maximum (maximum-a-posteriori, or MAP, approximation). The one-dimensional integral over  $b$  is carried out numerically. Given the MAP estimate of  $b$ , inferring a note  $n_k$  is done by finding the  $n_k$  that maximizes the

probability of generating the corresponding  $f_k$ . For the predictions, we used the MAP parameter estimates for  $b$  and  $q$  to compute predictive probability distributions for those frequencies that had not been used to make these estimates. We set  $b \in [100 \text{ Hz}, 1,000 \text{ Hz}]$  and  $\alpha = 2$ . The value of  $\alpha$  is not critical.

To determine a suitable value for the jitter SD  $\sigma$ , we recorded 10 human singers singing "Happy Birthday" two times each. Because this song is diatonic rather than purely harmonic, we used a song model assuming a chromatic scale, by replacing Eq. 2 with

$$f_k = b \cdot r^{n_k} \cdot \rho, \quad [4]$$

where  $r = \sqrt[12]{2}$  is the frequency ratio between two steps on a chromatic scale and  $n_k \geq 0$  an integer. The exact value of the SD of  $\log \rho$  is not critical but should be small enough so that neighboring steps on the scale do not mix owing to jitter. We chose 0.01. For each recording, we inferred  $b$  and the  $n_k$  as described above and computed the reconstructed frequency  $\hat{f}_k = b \cdot r^{n_k}$ . This allowed us to estimate the empirical SD of  $\log f_k - \log \hat{f}_k$  across all recordings and singers. We found a value of 0.014, which we subsequently used for  $\sigma$ .

The above generative model can be used not only to infer the base frequency and notes but also to determine how probable is an observed song under H0 versus H1. To avoid overfitting we used leave-one-out cross-validation (43), that is, we inferred  $b$  and  $q$  from all frequency observations in a song but one and then computed the predictive probability of the left-out observation.

**Validity Testing of Both Models on a Ground Truth Dataset.** To verify that both statistical approaches used here produce sensible results we used a ground truth dataset (GT). A ground truth is a dataset that shares relevant features with the data of interest but has controllable statistical properties. Our GT is a random sequence generator (technically, an exchangeable sequence generator) that can be continuously adjusted between producing perfectly harmonic sequences of frequencies and sequences with a nearly log-normal frequency distribution.

We generated these sequences from our Bayesian model with  $q$  set to the average hermit thrush song overtone distribution and  $b = 393.21$  Hz, which is the exponentiated average log-base frequency of the hermit thrush songs. The harmonicity of a generated sequence can be controlled by varying the jitter SD  $\eta$ . For very small  $\eta$ , the resulting frequency density is composed of distinct, evenly spaced peaks. This can be seen in Fig. 3 for  $\eta = 0.005$  and  $\eta = 0.015$ . Drawing a sequence from such a density yields a harmonic sequence, because virtually all probability is contained in the peaks. For larger  $\eta$ , the peaks begin to overlap until the harmonic structure disappears and the density looks log-normal ( $\eta = 0.3$  in Fig. 3). We drew sequences with  $K = 14$  notes, which corresponds to the rounded average song length in our sample of hermit thrush songs. To validate that our hypothesis comparison is indeed sensitive enough to separate sequences whose frequency distribution corresponds to an overtone series from sequences whose frequencies are derived from a log-normal frequency distribution, we drew 100 sequences for several values of  $\eta$  between 0.005 and 0.3, for a total of 1,100 sequences.

We also used the GT to determine the minimum number of notes  $K$  that a sequence (or song) should contain so that we have enough statistical power to determine whether the detected harmonicity of the sequence corresponds to the true harmonicity. Specifically, we searched for a  $K$  such that both  $p(\text{H1}|\text{harmonic song}) > 1 - 0.05 = 0.95$  and  $p(\text{H0}|\text{random song}) > 0.95$ . Given the value of 0.014 obtained for the empirical deviation  $\sigma$  with human singers, we treated sequences with a jitter of  $\eta \leq 0.01$  as clearly harmonic, and sequences with  $\eta \geq 0.03$  as clearly nonharmonic. From the GT, we computed the log-ratio that each note contributes on average in favor (or against) each hypothesis. We obtained a log-ratio of 0.290 for  $\eta = 0.01$  in favor of H0 and 1.015 for  $\eta = 0.01$  in favor of H1. Assuming that we have no initial preference for either H0 or H1, these values imply  $K = 2.9$  and  $K = 10.1$ , respectively. We therefore chose  $K = 10$  as the minimal number of notes for which we have enough statistical power to determine the harmonicity of a song or sequence. Note that using a slightly larger or smaller value of  $K$  did not significantly affect our results.

**ACKNOWLEDGMENTS.** We thank the Borror Laboratory of Bioacoustics, Kevin Colver, and Bernie Krause for recordings; Chris Hill and Sue Anne Zollinger for information on vocal anatomy; and Neeltje Boogert, Drew Rendall, W. Andrew Schloss, Ford Doolittle, Tacye Phillipson, Andrew Horn, and Neil Banas for constructive input. E.L.D. thanks Canada Council for the Arts, and W.T.F. and B.G. acknowledge the support of European Research Council Advanced Grant 230604. D.M.E. acknowledges support from the EU Commission, KoroBot FP7-ICT-2013-10/611909, Deutsche Forschungsgemeinschaft DFG GI 305/4-1 and Graduiertenkolleg-IRTG-1901-BrainAct, the Human Brain Project, and Medical Research Council Fellowship G0501319.

1. Harkleroad L (2006) *The Math Behind the Music* (Cambridge Univ Press, New York).
2. Carterette EC, Kendall R (1999) Comparative music perception and cognition. *The Psychology of Music*, ed Deutsch D (Academic, New York), pp 725–791.
3. Terhardt E (1984) The concept of musical consonance: A link between music and psychoacoustics. *Music Percept* 1(3):276–295.
4. Gill KZ, Purves D (2009) A biological rationale for musical scales. *PLoS ONE* 4(12):e8144.
5. Tierney AT, Russo FA, Patel AD (2011) The motor origins of human and avian song structure. *Proc Natl Acad Sci USA* 108(37):15510–15515.
6. Larsen ON, Goller F (1999) Role of syrinxal vibrations in bird vocalizations. *Proc R Soc Lond B Biol Sci* 266(1429):1609–1615.
7. Goller F, Larsen ON (2002) New perspectives on mechanisms of sound generation in songbirds. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 188(11–12):841–850.
8. Elemans CPH, Larsen ON, Hoffmann MR, van Leeuwen JL (2003) Quantitative modelling of the biomechanics of the avian syrinx. *Anim Biol* 53(2):183–193.
9. Mindlin GB, Laje R (2005) *The Physics of Birdsong* (Springer, New York).
10. Elemans CPH (2014) The singer and the song: The neuromechanics of avian sound production. *Curr Opin Neurobiol* 28C:172–178.
11. Darwin C (1871) *The Descent of Man and Selection in Relation to Sex* (John Murray, London).
12. Armstrong EA (1963) *A Study of Bird Song* (Oxford Univ Press, London).
13. Hartshorne C (1973) *Born to Sing: An Interpretation and World Survey of Bird Song* (Indiana Univ Press, Bloomington, IN).
14. Dobson CW, Lemon RE (1977) Birdsong as music. *J Acoust Soc Am* 61(3):888–890.
15. Araya-Salas M (2012) Is birdsong music? Evaluating harmonic intervals in songs of a Neotropical songbird. *Anim Behav* 84:309–313.
16. Thorpe WH, Hall-Craggs J, Hooker B, Hooker T, Hutchison R (1972) Duetting and antiphonal song in birds: Its extent and significance. *Behaviour Suppl* 18:1–193.
17. Doolittle E, Brumm H (2012) O Canto do Uirapuru: Consonant pitches and patterns in the song of the musician wren. *J Inter Mus Stud* 6(1):55–85.
18. Oldys H (1913) A remarkable hermit thrush song. *Auk* 30(4):538–541.
19. Wing AH (1951) Notes on the song series of a hermit thrush in the Yukon. *Auk* 68(2):189–193.
20. Mathews FS (1921) *Field Book of Wild Birds and their Music* (Applewood Books, Bedford, MA).
21. Ingraham SE (1938) Instinctive music. *Auk* 55(4):614–628.
22. Rivers JW, Kroodmsa DE (2000) Singing behavior of the hermit thrush. *J Field Ornithol* 71(3):467–471.
23. Roach SP, Johnson L, Phillmore LS (2012) Repertoire composition and singing behaviour in two eastern populations of the Hermit Thrush (*Catharus guttatus*). *Bioacoust* 21(3):239–252.
24. Benade AH (1990) *Fundamentals of Musical Acoustics* (Dover, New York).
25. Nowicki S, Marler P (1988) How do birds sing? *Music Percept* 5(4):391–426.
26. Nowicki S (1987) Vocal tract resonances in oscine bird sound production: Evidence from birdsongs in a helium atmosphere. *Nature* 325(6099):53–55.
27. Titze IR (1994) *Principles of Voice Production* (Prentice Hall, Englewood Cliffs, NJ).
28. Levin TC, Edgerton ME (1999) The throat singers of Tuva. *Sci Am* 281(3):80–87.
29. Lindestad PA, Södersten M, Merker B, Granqvist S (2001) Voice source characteristics in Mongolian “throat singing” studied with high-speed imaging technique, acoustic spectra, and inverse filtering. *J Voice* 15(1):78–85.
30. Soha JA, Marler P (2000) A species-specific acoustic cue for selective song learning in the white-crowned sparrow. *Anim Behav* 60(3):297–306.
31. Trehub SE (2000) Human processing predispositions and musical universals. *The Origins of Music*, eds Wallin NL, Merker B, Brown S (MIT Press, New York), pp 427–448.
32. Watanabe S, Uozumi M, Tanaka N (2005) Discrimination of consonance and dissonance in Java sparrows. *Behav Processes* 70(2):203–208.
33. Hulse SH, Bernard DJ, Braaten RF (1995) Auditory discrimination of chord-based spectral structures by European starlings (*Sturnus vulgaris*). *J Exp Psychol Gen* 124(4):409–423.
34. Porter D, Neuringer A (1984) Music discriminations by pigeons. *J Exp Psychol Anim Behav Process* 10(2):138–148.
35. Chiandetti C, Vallortigara G (2011) Chicks like consonant music. *Psychol Sci* 22(10):1270–1273.
36. Wright AA, Rivera JJ, Hulse SH, Shyan M, Neiwirth JJ (2000) Music perception and octave generalization in rhesus monkeys. *J Exp Psychol Gen* 129(3):291–307.
37. Cator LJ, Arthur BJ, Harrington LC, Hoy RR (2009) Harmonic convergence in the love songs of the dengue vector mosquito. *Science* 323(5917):1077–1079.
38. Schwartz DA, Howe CQ, Purves D (2003) The statistical structure of human speech sounds predicts musical universals. *J Neurosci* 23(18):7160–7168.
39. Ross D, Choi J, Purves D (2007) Musical intervals in speech. *Proc Natl Acad Sci USA* 104(23):9852–9857.
40. Patel AD, Iversen JR, Bregman MR, Schulz I (2009) Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Curr Biol* 19(10):827–830.
41. Schachner A, Brady TF, Pepperberg IM, Hauser MD (2009) Spontaneous motor entrainment to music in multiple vocal mimicking species. *Curr Biol* 19(10):831–836.
42. Boersma P, Weenink D (1992–2010) *Praat: Doing Phonetics by Computer* (Univ of Amsterdam), Version 5.1.29. Available at [www.praat.org/](http://www.praat.org/). Accessed June 18, 2012.
43. Bishop CM (2006) *Pattern Recognition and Machine Learning* (Springer, New York).