

Annual variation in vocal performance and its relationship with bill morphology in Lincoln's sparrows, *Melospiza lincolnii*

Keith W. Sockman*

Department of Biology and Curriculum in Neurobiology, University of North Carolina, Chapel Hill

ARTICLE INFO

Article history:

Received 12 September 2007

Initial acceptance 17 October 2007

Final acceptance 19 November 2008

Published online 8 January 2009

MS. number: A10855R4

Keywords:

annual variation
bill morphology
bird song
environmental change
Lincoln's sparrow
Melospiza lincolnii
sexual selection
signal production
trill performance

Morphology may affect behavioural performance through a direct, physical link or through indirect, secondary mechanisms. Although some evidence suggests that the bill morphology of songbirds directly constrains vocal performance, bill morphology may influence vocal performance through indirect mechanisms also, such as one in which morphology influences foraging and thus the ability to perform some types of vocal behaviour. This raises the possibility for ecologically induced variation in the relationship between morphology and behaviour. To investigate this, I used an information theoretic approach to examine the relationship between bill morphology and several measures of vocal performance in Lincoln's sparrows. I compared this relationship between two breeding seasons that differed markedly in ambient temperatures, phenology of habitat maturation and food abundance. I found a strong curvilinear relationship between bill shape (height/width) and vocal performance in the seemingly less hospitable season but not in the other, leading to a difference between seasons in the population's mean vocal performance. Currently, I do not know the cause of this annual variation. However, it could be due to the effects of bill shape on foraging and therefore on time budget, energy balance, or some other behavioural or physiological response that manifests mostly under difficult environmental conditions or, alternatively, to associations between male quality and both vocal performance and bill shape. Regardless of the cause, these results suggest the presence of an indirect, ecologically mediated link between morphology and behavioural performance, leading to annual variation in the prevailing environment of acoustic signals.

© 2008 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.

Behaviour is the product of nervous impulse, muscular contraction and anatomical movement. Individual differences in the morphology of structures involved in these components give rise to individual differences in behavioural performance (e.g. Bass et al. 1997). For instance, with all else equal, flexion will be slower for a large than for a small limb, simply because of the forces necessary to move limbs of different inertia. Although the mechanisms by which morphological variations cause differences in behavioural performance are often through this type of direct, physical linkage, in some cases the mechanistic linkage may be indirect and mediated by a complex series of intermediate steps involving everything from behavioural to physiological factors (Bass 1998). Understanding the consequences of morphological variation and the details of how it gives rise to behavioural differences is of critical importance because this variation can have implications for the evolution of behaviour.

In many songbird species, male song functions in signalling territorial intentions and in the attraction of females for mating

purposes (McGregor 1991). The singer's ability to influence receiver behaviour may depend on features of his song that reflect his quality, such as his ability to perform some types of song behaviour that are inherently difficult or costly (Gil & Gahr 2002). One example of song behaviour that is putatively difficult is the production of large changes in sound frequency (i.e. broad-frequency bandwidths) at a rapid pace. A singer modulates sound frequency through conformational change in at least a part of the vocal tract (Nowicki 1987; Fletcher & Tarnopolsky 1999; Riede et al. 2006), which includes the trachea, syrinx and bill. The degree that the vocal tract component moves should constrain the rate at which it completes the movement, and thus the completion of rapid, large movements that enable broad-frequency bandwidths should be more difficult than the completion of slow, large movements; rapid, small movements; or slow, small movements.

The trill components of many songbird species' vocalizations are consecutive repetitions of a single syllable type and may therefore provide a signal of an individual's ability to move the vocal tract broadly and rapidly. Podos (1997) found for several songbird species that as the rate of syllable production (trill rate) increases, maximal trill bandwidth decreases (also see Results, Fig. 1), leading to the hypothesis that trill bandwidth constrains trill rate (but see

* Correspondence: K. W. Sockman, Department of Biology, University of North Carolina, Chapel Hill, NC 27599-3280, U.S.A.

E-mail address: kws@unc.edu.

Cardoso et al. 2007) and leading to a definition of trill performance based on the simultaneous maximization of trill rate and frequency bandwidth (see *Methods*). Thus, high-performance trills may more effectively influence male (Illes et al. 2006) and female (Draganoiu et al. 2002; Ballentine et al. 2004) receivers than do low-performance trills, which are presumably less difficult to produce.

Because the bill is part of the songbird vocal tract, it follows that variation in bill morphology could give rise to variation in trill performance through a direct, physical linkage, at least to the extent that bill gape influences the sound frequency of a vocalization (Podos et al. 1995). A recent series of correlational studies are consistent with this hypothesis (Podos 2001; Podos & Nowicki 2004; Huber & Podos 2006). However, several physiological studies challenge it, at least in some species, because bill gape may not necessarily play a significant role in shaping sound frequency (Riede et al. 2006). Regardless of whether or not a direct, mechanical link between bill morphology and trill performance exists, indirect, secondary mechanisms may mediate an effect of bill morphology on trill performance, as alluded to above. Specifically, bill morphology can influence an individual's choice of or fit to its foraging niche (Price et al. 1984; Smith 1987; Grant & Grant 1995). This, in turn, may influence time budget, energy balance, or other behavioural or physiological responses that affect trills and other types of vocal performance, such as song length and song structure (e.g. Pflumm et al. 1984). This potential for an indirect, ecologically mediated relationship between bill morphology and vocal performance raises the prediction that this relationship changes as environmental conditions change. Specifically, under favourable environmental conditions, when most members of a population may have ample access to resources, vocal performance may be relatively high, regardless of bill morphology. Conversely, when conditions are limiting, bill morphology may play more of a role in resource acquisition and therefore vocal performance.

I tested this prediction in Lincoln's sparrows by examining annual variation in the relationship between bill morphology and vocal performance, using data from two breeding seasons that seemed to differ markedly in how favourable environmental conditions were to resource acquisition. Lincoln's sparrows breed in open-field habitats, which tend to favour the use of trills over other types of song vocalizations (Wiley & Richards 1982; Wiley 1991, 2009). Individual males (females do not sing) can sometimes insert up to four different trill types within a single 2 or 3 s song (Ammon 1995; Cicero & Benowitz-Fredericks 2000). If the differences in environmental conditions experienced over the 2 years of this study are relevant to an individual's behavioural output and if they differentially affect individuals based on their bill morphology, I would expect to observe a stronger relationship between bill morphology and vocal performance during the more limiting year than during the more favourable year.

METHODS

The U.S. Department of the Interior, Geological Service (permit 23370), the U.S. Fish and Wildlife Service (permit MB099926), the U.S. Department of Agriculture Forest Service (authorization COL258), the State of Colorado Department of Natural Resources Division of Wildlife (license 06TR1056A2), the Town of Silverton, Colorado, U.S.A., and the Institutional Animal Care and Use Committee of the University of North Carolina at Chapel Hill (protocol 05-138.0-A) each granted permission to conduct the procedures described in this study.

This study uses data collected during the 2005–2007 Lincoln's sparrow breeding seasons near Molas Pass, Colorado, U.S.A. (37.74°N, 107.69°W). At an elevation of 3250 m, the study site is an open-field,

wet meadow ca. 20 ha in area (Sockman 2008). Sub-alpine ecosystems of this type are particularly susceptible to interannual variation in climatic and biotic conditions (Morton 2002).

Environmental Data

Field assistants and I collected data on annual variation in habitat maturation, temperature and food availability. The goal of analysing these data was not to determine what environmental factors drive potential annual variation in the relationship between bill morphology and vocal performance but simply to examine how the years of primary interest differed in at least some aspects. Because I had song data from only 2005 and 2006, I include only 2005 and 2006 environmental data.

Because of rapid snow melt and the growth of new leaves, the low-lying, willow (*Salix glauca* and *Salix wolfii*) habitat in which these Lincoln's sparrows forage and nest matures rapidly over the course of their breeding season. To document gross seasonal and annual variation in habitat maturation, we captured digital photographs of the study site from the same single perspective each week throughout the breeding seasons (see *Results*, Fig. 2). After cropping each photograph, I used Spot software (version 4.5.9.8 for the Macintosh, Diagnostic Instruments, Inc., Sterling Heights, MI, U.S.A.) to convert the green component of each pixel into one of 256 shades of grey and to remove the remaining colour components (in this case the blue and red). I did the same for the blue and red components of each pixel, making three greyscale images from each photograph. Using ImageJ (version 1.38x for the Macintosh, National Institutes of Health, Bethesda, MD, U.S.A.), I then quantified the mean grey value of each of five sampling locations in each greyscale image. The sampling locations were within willow patches known to support Lincoln's sparrow nests and foraging. My index of habitat maturity was the mean grey value for the green component, which I divided by the sum of the mean grey values for the red, green and blue components, to control for variation in exposure between photographs. In an effort to validate this procedure, I collected under different lighting conditions a digital image of each of three different, green objects and of each of three different, non-green objects (one blue, one brown, one grey). I applied the same procedure as that described above, and, despite very low power, the results clearly indicated that the two categories of objects differed in their greenness (ANOVA: $F_{1,4} = 11.55$, $P = 0.03$), with no overlap in greenness between replicates of different categories (greenness values for three green objects: 0.535, 0.561, 0.407; for three non-green objects: 0.343, 0.349, 0.288).

We collected temperature data from 2 June to 20 July, spanning the period from the earliest nest initiation through nearly all nest completions. Two digital data loggers (Onset, Bourne, MA, U.S.A.) in 2005 and six in 2006 positioned representatively across the study site recorded temperature every 5–10 min throughout the study.

To assess food availability, we captured flying arthropods in six traps positioned representatively across the study site. Each trap was made from a yellow, plastic tub (ca. 35 × 25 × 20 cm) three-fourths filled with water, to which we added a small amount of liquid detergent to reduce surface tension. We positioned each trap approximately 1 m above ground in a plywood frame supported by four threaded rods driven into the soil. We collected arthropods and resupplied the traps with water and detergent weekly. We stored arthropods in 70% ethanol, before drying them at 60 °C to determine dry mass. Although Lincoln's sparrows on our site forage on a variety of items (including seeds and nonflying arthropods), I have on several occasions observed them carrying to their nestlings the species that we often collected in the traps. Therefore, this means of arthropod collection and quantification probably

reflects, to a reasonable extent, some of the variation in food availability that the sparrows experienced.

Behavioural and Morphological Data

We found nests by searching habitat and observing adult behaviour. To examine annual variation in nesting phenology, I estimated the date of clutch initiation in nests at different stages of the nesting cycle as follows: (1) for nests found during laying, I subtracted the number of eggs (assuming one laid per day) minus one from the discovery date; (2) for nests found during incubation, I subtracted the mean incubation period (calculated from nests of known incubation period) and the number of eggs minus one from the hatching date; (3) for nests found with nestlings, I subtracted the mean nestling period (calculated from nests of known nestling period), the mean incubation period and the mean clutch size minus one from the fledging date (see Sockman 2008).

We captured adults in mist nets and seed-baited traps and measured bill height and width at the centre of the nares. We also measured bill length from the distal edge of the nares to the tip of the bill but did not analyse it, because I later realized that this particular method of measuring bill length probably reflected variation in the diameter of the nares and the wear of the bill tip more than it reflected variation in the length of the bill's functional elements. To estimate body size, we measured tarsus (first-axis loading, 0.34), wing cord (0.57), ninth primary (0.53) and outer rectrix (0.53) lengths and statistically combined these into a single composite variable using the first-axis (proportion of variance explained, 0.52) factor scores of a principal components analysis (see Sockman & Schwabl 2000). We affixed a unique combination of coloured leg bands to each bird for individual identification during song recordings.

Using two short-shotgun microphones (Sennheiser ME-66/K6, Wedemark, Germany), each connected to one of two digital recorders (Marantz PMD 660 and 670, Mahwah, NJ, U.S.A.) set to record uncompressed files sampled at 44.1 kHz, we collected audio recordings of male song during the morning hours (from ca. pre-dawn until 1000 hours) daily. After eliminating noise below the frequency range of the Lincoln's sparrow song with a high-pass filter, I generated spectrograms in the program WildSpectra (<http://www.unc.edu/~rhwiley/wildspectra/index.html>) under a transform size of 128 points, a frequency resolution of 344 Hz, and a temporal resolution of 2.90 ms. Using an automated routine in WildSpectra that determines the high and low frequencies within 3 dB of the peak amplitude and the start and stop times of each syllable based on user-defined amplitude thresholds, I calculated the trill rate, the frequency bandwidths of each syllable, the duration of each syllable and of each gap between syllables of a trill, and the length of each song (excluding those in low-quality recordings). Other studies (e.g. Podos 1997) have used a lower threshold (–24 dB) for measuring frequency bandwidth. However, because of its base-10 logarithm scale, the –3 dB threshold I used resulted in bandwidth measurements (approximate range 0.5–3.0 kHz) only marginally lower than those previously reported for this species (approximate range 1.0–4.0 kHz) (Podos 1997). Regardless, my use of a high threshold for calculating frequency bandwidth should have no effect on the annual variation in the relationship between bill morphology and vocal performance, the phenomenon of primary interest here. I then plotted the trill's mean syllable bandwidth in kHz against the trill rate in Hz and calculated an upper-bound regression following Podos (1997). This is the estimate of the upper boundary (hypotenuse) of a right-triangular distribution and thus represents the theoretical limit to this performance measure (see Fig. 1). To determine the performance value of each trill, I then calculated the difference between each

trill's value for its location on this bivariate plot and the nearest point on the upper-bound regression: the smaller, more negative this difference, the further below the upper-bound regression and the lower the trill performance. In analyses, I used only the trill in each song with the highest trill performance measure. I also calculated an index of respiratory performance by dividing the duration of each syllable (except the last) in each trill by the total duration of the syllable plus the gap following it. This measurement is similar but not identical to another index of respiratory performance recently proposed (Forstmeier et al. 2002). I used the mean of this ratio for each trill and used only the trill in each song with the highest respiratory performance index. I also counted the syllables per song and phrases (clusters of the same syllable type) per song. The phrase type defined by a single syllable type was never repeated in an individual song. Therefore, phrase count is also a measure of an individual song's syllable complexity.

Analysis

As is typical of field studies, sample sizes varied greatly, resulting in numerous observations for which I had only partial information. In addition, the data set consisted of a hierarchically nested combination of fixed and random effects, each of which may differ from the others in its correlation structure (e.g. songs nested within audio recordings nested within individual males). Therefore, I used a mixed-model framework (using Stata IC 9.1 for the Macintosh, Stata Corporation, College Station, TX, U.S.A.), which is readily amenable to complex data sets such as this (Burton et al. 1998; Goldstein et al. 2002; Rabe-Hesketh & Skrondal 2005). For a thorough description of applying this mixed, multilevel modelling approach to field studies on measures of animal performance, see Sockman et al. (2008). Parameters were estimated using restricted maximum likelihood. I assessed model fit using an information theoretic approach, in which I calculated the Akaike information criterion (AIC) (Burnham & Anderson 2002) value for a null model containing only an intercept and for each model in which I added the various combinations of the predictors of interest and their interactions. A model's AIC value reflects its goodness of fit relative to its number of parameters, using the log-likelihood of the model while penalizing for each parameter. I then report the *P* values from the statistical hypothesis tests conducted for only those predictors in the particular model with the best (lowest) AIC value.

RESULTS

Relationship between Frequency Bandwidth and Trill Rate

I determined the mean frequency bandwidth and the trill rate for 1680 trills in 815 songs from 140 audio recordings of 37 males. The distribution of mean frequency bandwidth plotted against trill rate was relatively triangular, although one corner of the triangle, where trill rate was maximal and bandwidth minimal, seemed truncated, suggesting an upper boundary not only to trill performance but also to trill rate (Fig. 1). Importantly, the upper-bound regression tended downward ($R^2 = 0.578$, $P = 0.004$), indicating triangularity in the expected orientation (Podos 1997).

Annual Variation in Environmental Conditions

To determine whether the phenology of habitat maturation varied annually, I used a mixed model with observation nested within year as a random coefficient on the square of week and with year nested within sampling area as a random intercept. The AIC best model included week, the square of week, year and the interaction between week and year (each effect: $P < 0.001$) as

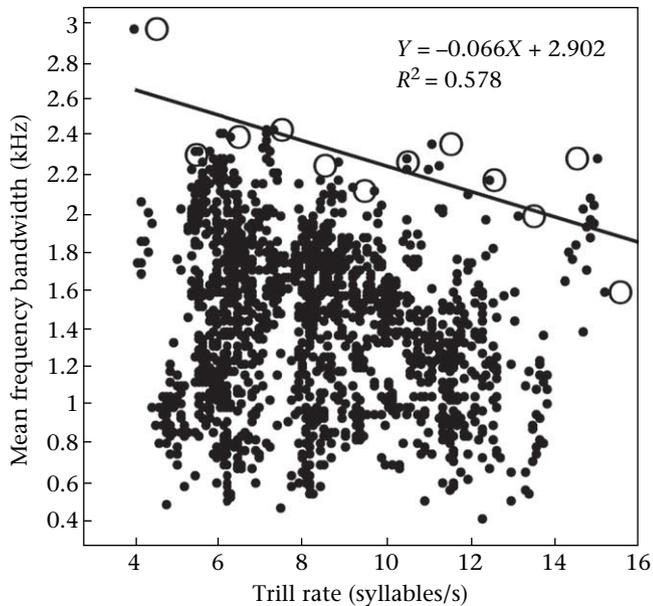


Figure 1. Trill performance in Lincoln's sparrows. The performance of individual trills is represented by the mean frequency bandwidth (-3 dB) for a given trill rate (solid symbols). The theoretical limit to trill performance is represented by an upper-bound regression (line) through the bandwidth maxima of each 1 Hz trill-rate bin (open symbols).

predictors. Examination of Fig. 2 together with these statistical results indicated that habitat maturation proceeded more rapidly in 2006 than in 2005.

I did not have accurate data on daily high temperatures. To determine whether daily low temperatures varied annually, I used a mixed model with observation nested within year as a random intercept and year nested within temperature logger as a random intercept. The AIC best model included date ($P < 0.001$) and year ($P = 0.07$) as predictors. The daily low showed a seasonal increase, regardless of year, and, on average, was lower in 2005 than in 2006 (Fig. 2). For daily means, I used a mixed model with observation nested within year as a random coefficient on date and with year nested within temperature logger as a random intercept. The AIC best model included date, year and their interaction (each effect: $P < 0.001$) as predictors. The interaction effect and inspection of Fig. 2 indicated that, in the first part of the breeding season, daily mean temperatures were lower in 2005 than in 2006, and in the second part of the breeding season, daily mean temperatures were higher in 2005 than in 2006.

To determine whether arthropod dry mass varied annually, I used a mixed model with observation nested within year as a random intercept and year nested within trap as a random intercept. The AIC best model included date ($P < 0.001$) and year ($P = 0.1$) as predictors. On average, arthropod dry mass was greater in 2006 than in 2005 (Fig. 2).

Perhaps the strongest evidence that environmental conditions in 2005 differed from those in 2006 comes from data on nesting phenology (Fig. 2). Nests in 2006 were initiated, on average, more than 1 week (17% of the range of nest initiation times) earlier in the breeding season than were nests in 2005 (ANOVA: $F_{1,126} = 33.01$, $P < 0.0001$).

Bill Morphology and Vocal Performance

I had song data for 13 males in 2005 and 17 in 2006. Four of the males in 2006 were also sampled in 2005, making for a total of 26 males across both years of the study. I examined annual variation in

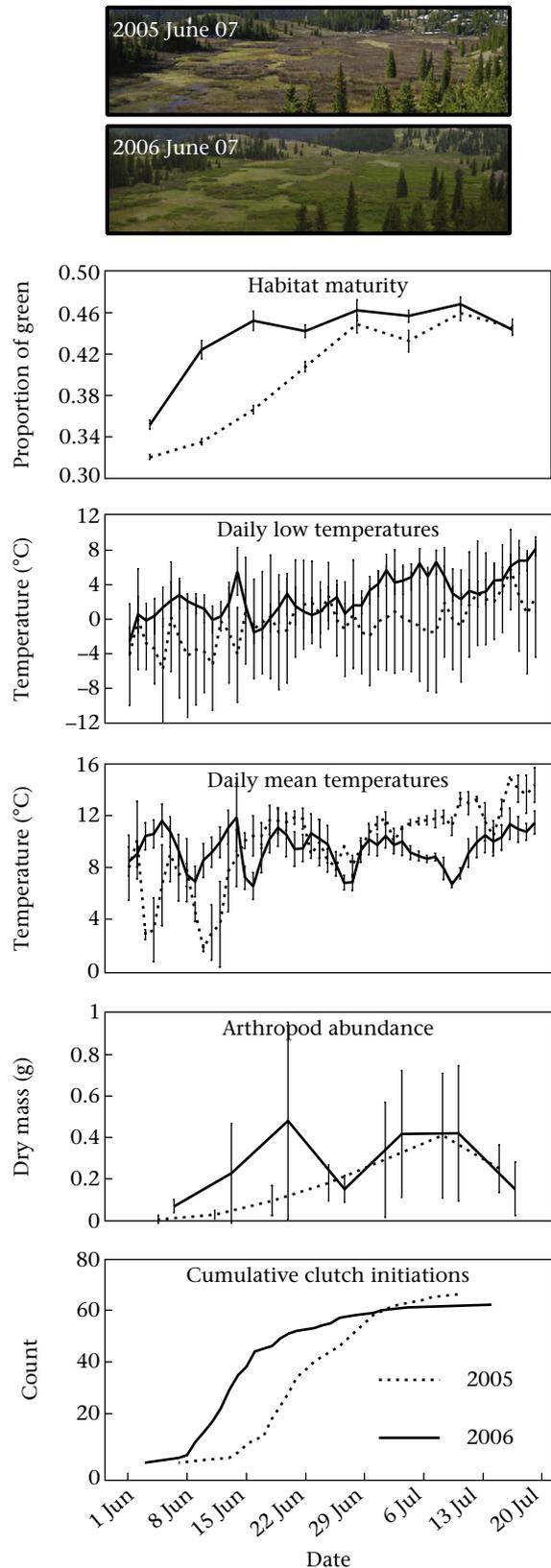


Figure 2. Annual variation in environmental conditions and in nesting phenology of Lincoln's sparrows at Molas Pass, Colorado, U.S.A. For habitat maturity, daily low temperature, daily mean temperature and arthropod abundance, lines are means with 95% confidence whiskers of replicate samples. Proportion of green is the proportion of green relative to green plus red plus blue in pixels of digital images of each sampling location at each time.

the relationship between bill morphology and trill performance using mixed models with song nested within audio recording as a random intercept and with audio recording nested within individual male as a random intercept (691 songs in 111 recordings of 26 males). The fact that individuals are sampled multiple times (multiple songs within multiple recordings across 2 years) is accounted for statistically by the hierarchically nested, mixed effects linear regression, which models the correlation structure of each nested random factor independently. Thus, each song or recording is not treated as independent relative to factors that differ between individuals, such as bill shape. As predictors, I used bill height, bill width or bill shape (height/width). I also tested the effects of the squares of these predictors, of their interactions with year, of year, and of body size, date and time of day. The AIC best model included bill shape, the square of bill shape, year, the interaction between bill shape and year, and the interaction between the square of bill shape and year (each effect: $P \leq 0.001$) as predictors (Table 1). Based on AIC values, this model was more efficient than models in which I substituted bill shape with height or width (and their squares and interactions) or in which I also included body size, date or time of day. There was also substantial intraindividual variation in trill performance, which was not efficiently explained by the intraindividual variables date and time of day. Additionally, I found no evidence from AIC values for an effect of year on bill shape.

It is possible that the effects of the set of predictors above on trill performance are due to their effects on one of the two components that go into calculating trill performance. Therefore, before conducting post hoc analyses from the preceding results on each year separately, I first examined the effects of these predictors on trill rate alone and on mean frequency bandwidth alone (each response: 691 songs in 111 recordings of 26 males), using the same hierarchical model as that described above. The AIC best model for each of these two responses had the same set of predictors as the AIC best model for trill performance (Table 1): bill shape (trill rate: $P = 0.001$; bandwidth: $P = 0.005$), the square of bill shape (trill rate: $P = 0.001$; bandwidth: $P = 0.006$), year (trill rate: $P = 0.002$; bandwidth: $P = 0.001$), the interaction between bill shape and year (trill rate: $P = 0.003$; bandwidth: $P = 0.002$), and the interaction between the square of bill shape and year (trill rate: $P = 0.003$; bandwidth: $P = 0.002$). These results raise the possibility that the effects of these predictors on trill performance might be due to their effects on trill rate alone or frequency bandwidth alone. So, in an attempt to acquire some additional understanding, I reanalysed trill rate as a response variable with the above set of predictors and added mean frequency bandwidth to the model as a covariate, under the assumption that a further improvement in model fit with the addition of the covariate would suggest that the effect of the predictors on trill performance is not best explained using only one component of trill performance. I also examined the analogous model with mean frequency bandwidth as the response and trill rate as a covariate. Indeed, in both cases, the addition of the covariate (each response: $P < 0.001$) improved the fit of the model, based on AIC values (Table 1), suggesting that variation in one component of trill performance is constrained, in part, by variation in the other. For simplicity, I hereafter analyse and discuss trill performance as opposed to the components that make up this variable.

For post hoc analyses, I examined the effects of bill shape and its square on trill performance for each year separately and found that the AIC best model for 2005 was one with bill shape and its square (each effect: $P = 0.01$) as predictors, whereas the AIC best model for 2006 was one with only an intercept. Examination of Fig. 3 together with these statistical results indicated that trill performance was, on average, higher in 2006 than in 2005, and that in 2005, trill

Table 1

Parameter estimates for AIC best models of vocal performance in Lincoln's sparrows

Response	Estimate	SE	P	AIC
Predictor				
Trill performance				
Intercept	-14.76	3.785	<0.001	-1237.8
Bill shape	23.78	6.337	<0.001	-1234.1
Bill shape ²	-9.702	2.645	<0.001	-1239.4
Year	13.48	3.745	<0.001	-1233.1
Bill shape × year	-21.76	6.225	<0.001	-1235.7
Bill shape ² × year	8.736	2.574	0.001	-1246.7
Trill rate (Hz)				
Intercept	-183.5	71.00	0.01	2591.2
Bill shape	309.5	119.0	0.009	2585.0
Bill shape ²	-124.8	49.64	0.01	2575.7
Year	160.9	66.77	0.02	2572.0
Bill shape × year	-262.4	110.8	0.02	2567.7
Bill shape ² × year	107.0	45.76	0.02	2551.8
Bandwidth	1.189	0.2177	<0.001	2526.0
Bandwidth (kHz)				
Intercept	-25.26	12.20	0.04	151.5
Bill shape	43.87	20.44	0.03	152.9
Bill shape ²	-18.00	8.526	0.04	148.3
Year	30.57	11.28	0.007	153.8
Bill shape × year	-49.53	18.72	0.008	151.0
Bill shape ² × year	19.88	7.729	0.01	139.5
Trill rate	0.03510	0.006434	<0.001	120.8
Song length (s)				
Intercept	-38.45	10.79	<0.001	-175.0
Bill shape	66.73	18.08	<0.001	-174.1
Bill shape ²	-27.19	7.549	<0.001	-179.7
Year	39.21	9.883	<0.001	-180.8
Bill shape × year	-63.34	16.40	<0.001	-184.7
Bill shape ² × year	25.50	6.769	<0.001	-201.7
Phrase count				
Intercept	-71.96	38.92	0.06	3180.3
Bill shape	128.3	65.32	0.05	3178.5
Bill shape ²	-52.66	27.31	0.05	3171.3
Year	78.43	27.70	0.005	3169.1
Bill shape × year	-129.0	45.52	0.005	3168.1
Bill shape ² × year	52.90	18.57	0.004	3154.5
Syllable count				
Intercept	-102.1	91.08	>0.2	6216.1
Bill shape	186.7	152.2	>0.2	6211.6
Bill shape ²	-73.76	63.40	>0.2	6203.7
Year	161.3	82.84	0.05	6195.4
Bill shape × year	-261.4	136.2	0.06	6192.2
Bill shape ² × year	106.3	55.62	0.06	6180.7
Song size				
Intercept	-201.3	44.60	<0.001	1976.8
Bill shape	329.3	74.72	<0.001	1973.6
Bill shape ²	-133.9	31.19	<0.001	1964.5
Year	193.8	41.82	<0.001	1960.1
Bill shape × year	-315.0	69.41	<0.001	1954.1
Bill shape ² × year	127.5	28.66	<0.001	1929.4

AIC values apply to models containing the predictor to its left and all predictors above it. For example, in modelling song length, the AIC value of -179.7 applies to a model containing an intercept, bill shape and the square of bill shape. Not shown but also assessed with AIC values were numerous additional models, such as those in which predictors were introduced in alternate sequences (e.g. year before bill shape) and in which additional predictors such as bill width, bill height, body size, time of day and date were assessed. See text for a summary of these assessments. Trill performance is the difference between the value of a trill's location on a bandwidth-trill rate bivariate plot and the nearest point on the upper-bound regression in Fig. 1, with smaller, more negative differences corresponding to lower performance. Song size is the first-axis factor score of a principal components analysis of song length, syllable count and phrase count. Bill shape is height/width. Year is coded 0 for 2005 and 1 for 2006.

performance increased with the vertical ovality of the bill (bills that were taller than they were wide) to a point, beyond which trill performance began to decline. However, in 2006, I found no obvious relationship between bill shape and trill performance.

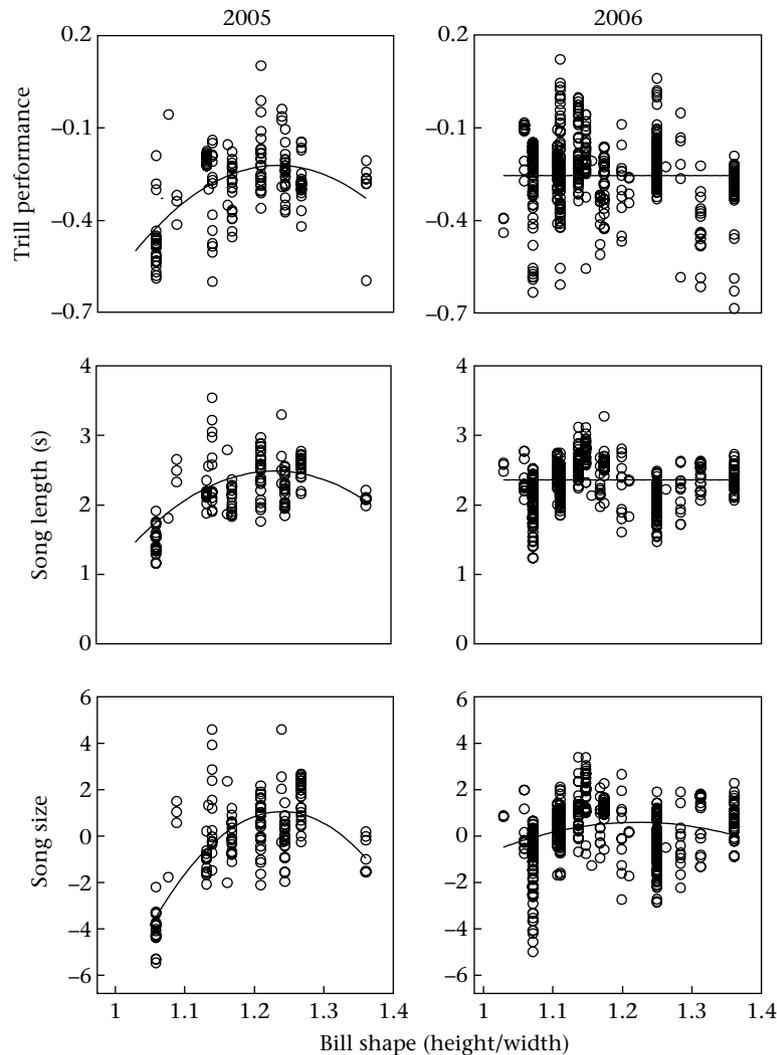


Figure 3. Annual variation in the relationship between bill shape and vocal performance in Lincoln's sparrows. Three measures of vocal performance (trill performance, song length, song size) are depicted, and each point in each graph represents the performance value (trill performance and song size: 691 points; song length: 756 points) of an individual song. Each column of points, aligned along a single bill shape value, corresponds to an individual male in the analyses and illustrates the substantial variation in vocal performance between an individual's songs. Trill performance is the difference between the value of a trill's location on a bandwidth–trill rate bivariate plot and the nearest point on the upper-bound regression in Fig. 1, with smaller, more negative differences corresponding to lower performance. Song size is the first-axis factor score of a principal components analysis of song length, syllable count and phrase count. Prediction lines are from post hoc AIC best models for each year separately.

I also examined annual variation in the effect of bill shape on respiratory performance (691 songs in 111 recordings of 26 males), song length (756 songs in 112 recordings of 26 males), syllable count (1465 songs in 167 recordings of 29 males) and phrase count (1466 songs in 167 recordings of 29 males). I modelled each of these responses using the same hierarchical structure that I used for modelling trill performance. The model that best described variation in respiratory performance was one with only a constant. Note, however, that my measure of respiratory performance differed from that used by Forstmeier et al. (2002), in that I quantified the duration of whole syllables and gaps between syllables as opposed to the duration of notes within syllables. This or other reasons, such as the possibility of low variation in my measure of respiratory performance, might account for this lack of effect, and, consequently, I do not discuss it further.

The AIC best model for each of the other responses (song length, syllable count, phrase count) had the same set of predictors as the AIC best for trill performance (Table 1): bill shape (song length: $P < 0.001$; syllable count: $P > 0.2$; phrase count: $P = 0.049$), the square of bill shape (song length: $P < 0.001$; syllable count: $P > 0.2$;

phrase count: $P = 0.054$), year (song length: $P < 0.001$; syllable count: $P = 0.05$; phrase count: $P = 0.005$), the interaction between bill shape and year (song length: $P < 0.001$; syllable count: $P = 0.06$; phrase count: $P < 0.005$) and the interaction between the square of bill shape and year (song length: $P < 0.001$; syllable count: $P = 0.06$; phrase count: $P = 0.004$). Moreover, for post hoc analysis of each of these response variables, the AIC best model for 2005 was always one with bill shape (song length: $P = 0.005$; syllable count: $P > 0.2$; phrase count: $P = 0.1$) and its square (song length: $P = 0.006$; syllable count: $P > 0.2$; phrase count: $P = 0.1$), whereas that for 2006 was always one with only an intercept. These statistical results together with examination of Fig. 3 indicated that trill performance, song length, syllable count (not shown in Fig. 3) and phrase count (not shown in Fig. 3) were, on average, higher in 2006 than in 2005, and that in 2005, each of these song measures increased with the vertical ovality of the bill up to a point, after which they began to decrease. In 2006, I found no obvious relationship between these performance measures and bill shape.

Song length, syllable count and phrase count are all conceptually related, positively correlated with each other (each pairwise

comparison: $P < 0.0001$), and thus should not necessarily be treated as independent responses. Therefore, I used the first-axis (proportion of variance explained, 0.75) factor scores of a principal components analysis on song length (first-axis loading, 0.64), syllable count (0.57) and phrase count (0.51) to construct a single, composite response variable of performance that I refer to as song size. Again, the AIC best model in the analysis of song size, modelled with the same hierarchical structure as those above (756 songs in 112 recordings of 26 males), was one having each of the same predictors as those of the above models (each effect: $P < 0.001$) (Table 1). For post hoc analysis, the AIC best model for both 2005 and 2006 was one with bill shape (2005: $P < 0.001$; 2006: $P > 0.2$) and its square (2005: $P = 0.002$; 2006: $P > 0.2$). These results and examination of Fig. 3 indicate that this measure of song size was, on average, higher in 2006 than in 2005, and that in 2005 and less so in 2006, song size increased with the vertical ovality of the bill up to a point, after which it began to decrease. Note that trill performance also positively correlated with song length, syllable count and phrase count (each pairwise comparison: $P < 0.0001$).

DISCUSSION

The relationship between bill morphology and vocal performance in male Lincoln's sparrows differed between 2 consecutive years that themselves differed markedly in their environmental conditions. In the year (2005) with cooler initial ambient temperatures, delayed habitat maturation and less abundant flying arthropods and, thus, when conditions seemed less hospitable, the relationship with trill performance was strongly curvilinear, with maximal performance corresponding to an intermediate bill shape and lower performance associated with shapes that were more extreme. However, when conditions seemed more hospitable (in 2006), I found no such relationship between bill shape and trill performance. Additionally, a very similar interaction between year and bill morphology applied to several other measures of vocal performance. Together these results suggest the presence of an indirect, ecologically mediated link between bill morphology and vocal performance that gave rise to a difference between years in the mean vocal performance of the population.

The conclusion that the relationship between vocal performance and bill morphology differs between years rests heavily on the statistical support for this difference. Although P values are often helpful in the interpretation of data, importantly, my overall conclusion does not depend on the dubious comparison of P values (Gelman & Stern 2006), some of which may be low (e.g. effect of bill shape on trill performance in 2005) and some high (e.g. effect of bill shape on trill performance in 2006). Instead, I took an information theoretic approach to examine the fit of various models using a data set assumed to be correct. This approach is more intuitive than the tradition of assuming a null hypothesis is correct and then determining the probability of a data set (Gibbons et al. 2007), and it is not encumbered by the arbitrariness of choosing a nominal alpha (e.g. 0.05). Interpreting how, exactly, the years differed in the relationship between bill morphology and vocal performance requires some speculation. Still, even without knowing how the 2 years differed in this relationship, the fact that they differed at all is of primary importance in this study and is supported by information theory and the statistical reliability of interactions between year and bill shape and between year and the square of bill shape.

What might the annual difference in the relationship between bill shape and vocal performance suggest in terms of how morphological variation can influence behavioural performance? If my measure of bill shape does have a causal effect on vocal performance, it must be through indirect mechanisms or in conjunction with other factors, otherwise it is difficult to imagine

how this effect might operate in one year but not another or how it might affect such song measures as syllable count, phrase count and song length. In terms of an indirect mechanism for such a constraint, one possibility is that bill shape influences foraging performance or foraging niche when preferred forage is limited. When preferred forage is low in availability (e.g. 2005), birds might shift to an alternative that individuals with moderately shaped bills more easily harvest than individuals with extreme bill shapes do, potentially producing individual differences in time budget, energy balance, or other physiological or behavioural responses based on the interaction between bill shape and ecological conditions. Differences in time allocation, energy, or other aspects of behaviour or physiology may then enable some individuals to produce high-performance song while others cannot (see Pflumm et al. 1984). For example, if individuals with extreme-shaped bills require more time foraging to maintain some level of condition, they would have less time to spend singing (Gil & Gahr 2002) and practicing trills and therefore may be less adept at high-performance trills than would individuals with intermediate-shaped bills.

Alternatively, high-performance trills or long songs with many syllables or phrases may be more energetically costly than low-performance trills or short songs with few syllables or phrases. Therefore, high-performance trills or long songs may be more affordable to individuals that can harvest food resources better than other individuals. A long history of research on the role of bill morphology on foraging efficiency (Smith 1987; Benkman & Miller 1996; Herrel et al. 2005) and the direct and indirect energetic costs of song (Oberweger & Goller 2001; Thomas et al. 2003; Ward et al. 2003; Ward & Slater 2005b), particularly in cold, windy environments (Ward & Slater 2005a) such as that of the Lincoln's sparrow, lends some credibility to the hypothesis that food or energy mediates an effect of bill shape on vocal performance. Although I am not aware of evidence that high-performance trills or long songs are more energetically costly than low-performance trills or short songs are, Pflumm et al. (1984) found in yellow-bellied sunbirds, *Nectarinia venusta*, that experimental elevation of sugar concentration in nectar forage reduces the occurrence of certain syllable types and increases the song rate and the number of syllables in trills, suggesting a role for the energy content of food in song structure. These investigators also report that variation in the sunbird's song depends on the availability of insects. Because there is some evidence for an energetic constraint on song complexity (Garamszegi et al. 2006), this indirect mechanism might also explain variation in phrase count, which is a measure of within-song syllable complexity. Nevertheless, at this point, this mechanism is only a hypothesis. Annual variation in the environmental parameters that I measured and the relationship between bill morphology and vocal performance may have no causal basis, direct or indirect.

There is also some evidence from other studies for an alternative possibility, that both variation in bill shape and variation in vocal performance under limiting conditions are products of individual differences in developmental experience. Experimental reduction in the number of nestmates during rearing increases bill height in adult zebra finches, *Taeniopygia guttata* (de Kogel 1997). With fewer nestmates competing for food, these individuals presumably have a more favourable developmental experience, which is reflected later in their adult bill morphology. It is already becoming fairly well established in several songbird species, including the zebra finch, that developmental experience can affect the quality of adult song in terms of its utility in attracting a mate (Nowicki et al. 1998, 2000, 2002; Buchanan 2000; Buchanan et al. 2003; Spencer et al. 2003, 2004, 2005; Nowicki & Searcy 2005; but see Gil et al. 2006). Perhaps bill shape and vocal performance during difficult conditions are both incidental consequences of developmental

conditions in Lincoln's sparrows. But again, this hypothesis has yet to be tested.

Results from this study raise several questions about bill shape from both the ultimate and proximate perspectives. For instance, why does bill shape vary enough between individuals to reveal this relationship? Do extreme bill shapes provide some advantage that offsets disadvantages in vocal performance in some years? Are intermediate-shaped bills difficult to develop and therefore limited to only a subset of individuals? These results also raise questions about the mechanism by which the songs of individuals with extreme bill shapes change. Is variation in songs due to plasticity in song production, whereby impoverished individuals facultatively narrow bandwidth, slow trill rate, reduce the number of syllables or phrases, or shorten songs (see Pflumm et al. 1984)? Or, instead, is variation in performance due to the use of particular songs (i.e. facultatively chosen from a large repertoire; Cicero & Benowitz-Fredericks 2000) that tend to be cheaper because the trills within them tend to have narrower bandwidth or slower trill rates, or because the songs are shorter, with fewer syllables or fewer phrases? Economic modulation of song effort based on the likely benefits of attracting a mate has been shown previously in Cassin's finches, *Carpodacus cassinii* (Sockman et al. 2005). In the above-mentioned scenarios, economic modulation would be based on variation not necessarily in benefits but rather in costs.

From the perspective of signal receivers, it is interesting that individuals with extreme-shaped bills varied annually in their mean vocal performance, causing the population mean to vary annually. Because receivers may assess the quality of a signal by comparing it to others in the population (Wiegmann et al. 1996; Jennions & Petrie 1997), receivers would need to exercise some annual flexibility in the thresholds they set for high signal quality when the mean signalling environment changes (Sockman 2007). In many songbird species, females choose mates based, in part, on several aspects of male song (McGregor 1991; Searcy 1992; Searcy & Yasukawa 1996; Gil & Gahr 2002), including trill performance (Draganoiu et al. 2002; Ballentine et al. 2004). Because mean vocal performance was highly variable between individuals in 2005 but not in 2006 and because mean performance in 2005 differed from that in 2006, the information content of this signal differed between these years. If Lincoln's sparrows use mate choice criteria similar to those in many other species and show plasticity in their threshold for quality, I would predict a difference between 2005 and 2006 in the mate choice patterns in this population.

Acknowledgments

I thank E. S. Bakunova, A. M. Byerly, C. R. Campbell, M. B. Danforth, A. M. Mascarenhas, D. M. Racke, S. E. Schulwitz, C. A. Schutte, K. L. Supler, S. R. Vora, C. M. Weimer and B. A. Whitman for their assistance in data collection. I also thank R. H. Wiley for customizing his WildSpectra software to better suit the idiosyncrasies of my sound analyses, J. D. Sockman for help with the analytical geometry and J. Podos for comments on a previous draft of this manuscript. A grant from the National Institutes of Health (NIH R01 grant number NS055125) and a U.N.C. Award from the R.J. Reynolds fund to K. W. S. provided funding.

References

Ammon, E. M. 1995. Lincoln's sparrow (*Melospiza lincolni*). In: *The Birds of North America Online* (Ed. by A. Poole). Ithaca, New York: Cornell Lab of Ornithology. <<http://bna.birds.cornell.edu/bna/species/191>> doi:10.2173/bna.191.

Ballentine, B., Hyman, J. & Nowicki, S. 2004. Vocal performance influences female response to male bird song: an experimental test. *Behavioral Ecology*, **15**, 163–168.

Bass, A. H. 1998. Behavioral and evolutionary neurobiology: a pluralistic approach. *American Zoologist*, **38**, 97–107.

Bass, A. H., Bodnar, D. A. & McKibben, J. R. 1997. From neurons to behavior: vocal-acoustic communication in teleost fish. *Biological Bulletin*, **192**, 158–160.

Benkman, C. W. & Miller, R. E. 1996. Morphological evolution in response to fluctuating selection. *Evolution*, **50**, 2499–2504.

Buchanan, K. L. 2000. Stress and the evolution of condition-dependent signals. *Trends in Ecology & Evolution*, **15**, 156–160.

Buchanan, K. L., Spencer, K. A., Goldsmith, A. R. & Catchpole, C. K. 2003. Song as an honest signal of past developmental stress in the European starling (*Sturnus vulgaris*). *Proceedings of the Royal Society of London, Series B*, **270**, 1149–1156.

Burnham, H. P. & Anderson, D. R. 2002. *Model Selection and Multimodel Inference*. New York: Springer-Verlag.

Burton, P., Gurrin, L. & Sly, P. 1998. Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level mixed modeling. *Statistics in Medicine*, **17**, 1261–1291.

Cardoso, G. C., Atwell, J. W., Ketterson, E. D. & Price, T. D. 2007. Inferring performance in the songs of dark-eyed juncos (*Junco hyemalis*). *Behavioral Ecology*, **18**, 1051–1057.

Cicero, C. & Benowitz-Fredericks, Z. M. 2000. Song types and variation in insular populations of Lincoln's sparrow (*Melospiza lincolni*), and comparisons with other *Melospiza*. *Auk*, **117**, 52–64.

Draganoiu, T. I., Nagle, L. & Kreutzer, M. 2002. Directional female preference for an exaggerated male trait in canary (*Serinus canaria*) song. *Proceedings of the Royal Society of London, Series B*, **269**, 2525–2531.

Fletcher, N. H. & Tarnopolsky, A. 1999. Acoustics of the avian vocal tract. *Journal of the Acoustical Society of America*, **105**, 35–49.

Forstmeier, W., Kempenaers, B., Meyer, A. & Leisler, B. 2002. A novel song parameter correlates with extra-pair paternity and reflects male longevity. *Proceedings of the Royal Society of London, Series B*, **269**, 1479–1485.

Garamszegi, L. Z., Moreno, J. & Møller, A. P. 2006. Avian song complexity is associated with high field metabolic rate. *Evolutionary Ecology Research*, **8**, 75–90.

Gelman, A. & Stern, H. 2006. The difference between 'significant' and 'not significant' is not itself statistically significant. *American Statistician*, **60**, 1–4.

Gibbons, J. M., Crout, N. M. J. & Healey, J. R. 2007. What role should null-hypothesis significance tests have in statistical education and hypothesis falsification? *Trends in Ecology & Evolution*, **22**, 445–446.

Gil, D. & Gahr, M. 2002. The honesty of bird song: multiple constraints for multiple traits. *Trends in Ecology & Evolution*, **17**, 133–141.

Gil, D., Naguib, M., Riebel, K., Rutstein, A. & Gahr, M. 2006. Early condition, song learning, and the volume of song brain nuclei in the zebra finch (*Taeniopygia guttata*). *Journal of Neurobiology*, **66**, 1602–1612.

Goldstein, H., Browne, W. & Rasbash, J. 2002. Partitioning variation in multilevel models. *Understanding Statistics*, **1**, 223–231.

Grant, P. R. & Grant, B. R. 1995. Predicting microevolutionary responses to directional selection on heritable variation. *Evolution*, **49**, 241–251.

Herrel, A., Podos, J., Huber, S. K. & Hendry, A. P. 2005. Bite performance and morphology in a population of Darwin's finches: implications for the evolution of beak shape. *Functional Ecology*, **19**, 43–48.

Huber, S. K. & Podos, J. 2006. Beak morphology and song features covary in a population of Darwin's finches (*Geospiza fortis*). *Biological Journal of the Linnean Society*, **88**, 489–498.

Illes, A. E., Hall, M. L. & Vehrencamp, S. L. 2006. Vocal performance influences male receiver response in the banded wren. *Proceedings of the Royal Society of London, Series B*, **273**, 1907–1912.

Jennions, M. D. & Petrie, M. 1997. Variation in mate choice and mating preferences: a review of causes and consequences. *Biological Reviews*, **72**, 283–327.

de Kogel, C. H. 1997. Long-term effects of brood size manipulation on morphological development and sex-specific mortality of offspring. *Journal of Animal Ecology*, **66**, 167–178.

McGregor, P. K. 1991. The singer and the song: on the receiving end of bird song. *Biological Reviews*, **66**, 57–81.

Morton, M. L. 2002. The mountain white-crowned sparrow: migration and reproduction at high altitude. *Studies in Avian Biology*, **24**, 1–236.

Nowicki, S. 1987. Vocal tract resonances in oscine bird song production: evidence from birdsongs in a helium atmosphere. *Nature*, **325**, 53–55.

Nowicki, S. & Searcy, W. A. 2005. Song and mate choice in birds: how the development of behavior helps us understand function. *Auk*, **122**, 1–14.

Nowicki, S., Peters, S. & Podos, J. 1998. Song learning, early nutrition and sexual selection in songbirds. *American Zoologist*, **38**, 179–190.

Nowicki, S., Hasselquist, D., Bensch, S. & Peters, S. 2000. Nestling growth and song repertoire size in great reed warblers: evidence for song learning as an indicator mechanism in mate choice. *Proceedings of the Royal Society of London, Series B*, **267**, 2419–2424.

Nowicki, S., Searcy, W. A. & Peters, S. 2002. Quality of song learning affects female response to male bird song. *Proceedings of the Royal Society of London, Series B*, **269**, 1949–1954.

Oberweger, K. & Goller, F. 2001. The metabolic cost of birdsong production. *Journal of Experimental Biology*, **204**, 3379–3388.

Pflumm, W., Comtesse, H. & Wilhelm, K. 1984. Sugar concentration and the structure of the sunbird's song. *Behavioral Ecology and Sociobiology*, **15**, 257–261.

Podos, J. 1997. A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). *Evolution*, **51**, 537–551.

Podos, J. 2001. Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature*, **409**, 185–188.

Podos, J. & Nowicki, S. 2004. Performance limits on birdsong production. In: *Nature's Music: The Science of Bird Song* (Ed. by P. Marler & H. Slabbekorn), pp. 318–342. New York: Academic Press.

- Podos, J., Sherer, J. K., Peters, S. & Nowicki, S.** 1995. Ontogeny of vocal tract movements during song production in song sparrows. *Animal Behaviour*, **50**, 1287–1296.
- Price, T. D., Grant, P. R., Gibbs, H. L. & Boag, P. T.** 1984. Recurrent patterns of natural selection in a population of Darwin's finches. *Nature*, **309**, 787–789.
- Rabe-Hesketh, S. & Skrondal, A.** 2005. *Multilevel and Longitudinal Modeling Using Stata*. College Station, Texas: Stata Press.
- Riede, T., Suthers, R. A., Fletcher, N. H. & Blevins, W. E.** 2006. Songbirds tune their vocal tract to the fundamental frequency of their song. *Proceedings of the National Academy of Sciences, U.S.A.*, **103**, 5543–5548.
- Searcy, W. A.** 1992. Song repertoire and mate choice in birds. *American Zoologist*, **32**, 71–80.
- Searcy, W. A. & Yasukawa, K.** 1996. Song and female choice. In: *Ecology and Evolution of Acoustic Communication in Birds* (Ed. by D. E. Kroodsma & E. H. Miller), pp. 454–473. Ithaca, New York: Cornell University Press.
- Smith, T. B.** 1987. Bill size polymorphism and intraspecific niche utilization in an African finch. *Nature*, **329**, 717–719.
- Sockman, K. W.** 2007. Neural orchestration of mate-choice plasticity in songbirds. *Journal of Ornithology*, **148**, S225–S230.
- Sockman, K. W.** 2008. Ovulation order mediates a trade-off between pre-hatching and post-hatching viability in an altricial bird. *Plos One*, **3**, e1785.
- Sockman, K. W. & Schwabl, H.** 2000. Yolk androgens reduce offspring survival. *Proceedings of the Royal Society of London, Series B*, **267**, 1451–1456.
- Sockman, K. W., Sewall, K. B., Ball, G. F. & Hahn, T. P.** 2005. Economy of mate attraction in the Cassin's finch. *Biology Letters*, **1**, 34–37.
- Sockman, K. W., Weiss, J., Webster, M. S., Talbott, V. & Schwabl, H.** 2008. Sex-specific effects of yolk-androgens on growth of nestling American kestrels. *Behavioral Ecology and Sociobiology*, **62**, 617–625.
- Spencer, K. A., Buchanan, K. L., Goldsmith, A. R. & Catchpole, C. K.** 2003. Song as an honest signal of developmental stress in the zebra finch (*Taeniopygia guttata*). *Hormones and Behavior*, **44**, 132–139.
- Spencer, K. A., Buchanan, K. L., Goldsmith, A. R. & Catchpole, C. K.** 2004. Developmental stress, social rank and song complexity in the European starling (*Sturnus vulgaris*). *Proceedings of the Royal Society of London, Series B, Supplement*, **271**, S121–S123.
- Spencer, K. A., Wimpenny, J. H., Buchanan, K. L., Lovell, P. G., Goldsmith, A. R. & Catchpole, C. K.** 2005. Developmental stress affects the attractiveness of male song and female choice in the zebra finch (*Taeniopygia guttata*). *Behavioral Ecology and Sociobiology*, **58**, 423–428.
- Thomas, R. J., Cuthill, I. C., Goldsmith, A. R., Cosgrove, D. F., Lidgate, H. C. & Proctor, S. L. B.** 2003. The trade-off between singing and mass gain in a daytime-singing bird, the European robin. *Behaviour*, **140**, 387–404.
- Ward, S. & Slater, P. J. B.** 2005a. Raised thermoregulatory costs at exposed song posts increase the energetic cost of singing for willow warblers, *Phylloscopus trochilus*. *Journal of Avian Biology*, **36**, 280–286.
- Ward, S. & Slater, P. J. B.** 2005b. Heat transfer and the energetic cost of singing by canaries, *Serinus canaria*. *Journal of Comparative Physiology A*, **191**, 953–964.
- Ward, S., Speakman, J. R. & Slater, P. J. B.** 2003. The energy cost of song in the canary, *Serinus canaria*. *Animal Behaviour*, **66**, 893–902.
- Wiegmann, D. D., Real, L. A., Capone, T. A. & Ellner, S.** 1996. Some distinguishing features of models of search behavior and mate choice. *American Naturalist*, **147**, 188–204.
- Wiley, R. H.** 1991. Associations of song properties with habitats for territorial oscine birds of eastern North America. *American Naturalist*, **138**, 973–993.
- Wiley, R. H.** 2009. Signal transmission in natural environments. In: *New Encyclopedia of Neuroscience* (Ed. by L. R. Squire), pp. 827–832. Oxford: Elsevier.
- Wiley, R. H. & Richards, D. G.** 1982. Adaptations for acoustic communication in birds: sound transmission and signal detection. In: *Acoustic Communication in Birds* (Ed. by D. E. Kroodsma, E. H. Miller & H. Ouellet), pp. 131–181. New York: Academic Press.