

László Z. Garamszegi · Thorsten J. S. Balsby ·
Ben D. Bell · Marta Borowiec · Bruce E. Byers ·
Tudor Draganoiu · Marcel Eens ·
Wolfgang Forstmeier · Paolo Galeotti · Diego Gil ·
Leen Gorissen · Poul Hansen · Helene M. Lampe ·
Stefan Leitner · Jan Lontkowski · Laurent Nagle ·
Erwin Nemeth · Rianne Pinxten · Jean-Marc Rossi ·
Nicola Saino · Aurélie Tanvez · Russell Titus ·
János Török · Els Van Duyse · Anders P. Møller

Estimating the complexity of bird song by using capture-recapture approaches from community ecology

Received: 17 March 2004 / Revised: 11 October 2004 / Accepted: 14 October 2004 / Published online: 2 December 2004
© Springer-Verlag 2004

Communicated by A. Cockburn

L. Z. Garamszegi (✉) · M. Eens · L. Gorissen · R. Pinxten ·
E. Van Duyse
Department of Biology, University of Antwerp,
Campus Drie Eiken, Universiteitsplein 1, 2610 Antwerp, Belgium
e-mail: laszlo.garamszegi@ua.ac.be
Tel.: +32-3-8202286
Fax: +32-3-8202271

L. Z. Garamszegi · J. Török
Department of Systematic Zoology and Ecology,
Eötvös Loránd University, Budapest, Hungary

T. J. S. Balsby
Department of Animal Behaviour,
University of Copenhagen, Copenhagen, Denmark

T. J. S. Balsby
Cornell Laboratory of Ornithology, Ithaca, New York, USA

B. D. Bell
School of Biological Sciences,
Victoria University of Wellington, Wellington, New Zealand

M. Borowiec
Department of Avian Ecology, Zoological Institute,
University of Wrocław, Wrocław, Poland

B. E. Byers
Department of Biology,
University of Massachusetts, Amherst, Mass., USA

T. Draganoiu · L. Nagle · A. Tanvez
Laboratoire d'Éthologie et de Cognition Comparées,
Université de Paris X-Nanterre, Nanterre, France

W. Forstmeier
Department of Animal and Plant Sciences,
University of Sheffield, Sheffield, UK

P. Galeotti
Laboratorio di Eco-Etologia—Dipartimento di Biologia Animale,
Università di Pavia, Pavia, Italy

D. Gil
Departamento de Ecología Evolutiva,
Museo Nacional de Ciencias Naturales (CSIC), Madrid, Spain

P. Hansen
Natural History Museum, Århus C, Denmark

H. M. Lampe
Department of Biology, Division of Zoology,
University of Oslo, Norway

S. Leitner
Max-Planck-Institut für Verhaltensphysiologie,
Seewiesen, Germany

J. Lontkowski
Museum of Natural History,
Wrocław University, Wrocław, Poland

E. Nemeth
Konrad Lorenz Institute for Comparative Ethology,
Vienna, Austria

J.-M. Rossi
Laboratoire d'Écologie,
Université Pierre et Marie Curie, Paris, France

N. Saino
Dipartimento di Biologia, Sez. Zoologia Scienze Naturali,
Università di Milano, Milano, Italy

R. Titus
45 Broadway, Stoughton, Wisconsin, USA

A. P. Møller
Laboratoire de Parasitologie Evolutive,
Université Pierre et Marie Curie, Paris, France

Abstract Repertoire size, the number of unique song or syllable types in the repertoire, is a widely used measure of song complexity in birds, but it is difficult to calculate this exactly in species with large repertoires. A new method of repertoire size estimation applies species richness estimation procedures from community ecology, but such capture-recapture approaches have not been much tested. Here, we establish standardized sampling schemes and estimation procedures using capture-recapture models for syllable repertoires from 18 bird species, and suggest how these may be used to tackle problems of repertoire estimation. Different models, with different assumptions regarding the heterogeneity of the use of syllable types, performed best for different species with different song organizations. For most species, models assuming heterogeneous probability of occurrence of syllables (so-called detection probability) were selected due to the presence of both rare and frequent syllables. Capture-recapture estimates of syllable repertoire size from our small sample did not differ significantly from previous estimates using larger samples of count data. However, the enumeration of syllables in 15 songs yielded significantly lower estimates than previous reports. Hence, heterogeneity in detection probability of syllables should be addressed when estimating repertoire size. This is neglected using simple enumeration procedures, but is taken into account when repertoire size is estimated by appropriate capture-recapture models adjusted for species-specific song organization characteristics. We suggest that such approaches, in combination with standardized sampling, should be applied in species with potentially large repertoire size. On the other hand, in species with small repertoire size and homogenous syllable usage, enumerations may be satisfactory. Although researchers often use repertoire size as a measure of song complexity, listeners to songs are unlikely to count entire repertoires and they may rely on other cues, such as syllable detection probability.

Keywords Bird song · Capture-recapture approach · Detection probability · Repertoire size · Species richness

Introduction

Bird species differ enormously in the variety of their songs, and diversity and versatility are key features of singing in passerines (Catchpole and Slater 1995). However, the quantitative characterization of the extreme variability of this temporal and structural vocal organization is a challenging task that makes comparisons among individuals and species difficult (Krebs and Kroodsma 1980; Kroodsma 1982). Repertoire size is a widely used measure of song complexity. It describes the number of unique vocal units, such as song or syllable types, that individuals can possess (Kroodsma 1982; Gil and Gahr 2002). A typical problem is that it is impossible to measure repertoire size explicitly in birds with virtuoso performance with large repertoires (e.g. Kroodsma and Parker 1977). Hence, in such cases instead of simple

enumeration, repertoire size must usually be estimated from a fraction of the whole repertoire (Catchpole and Slater 1995; Garamszegi et al. 2002a).

Additional difficulties may arise when species and even individuals differ in their singing styles. For instance, similar repertoire size may cover totally different song composition and virtuosity in different birds. On the other hand, the total amount of song material that two birds produce may be remarkably similar, but dramatic differences may be detected in repertoire size, because of the specific organization of the song output.

It has been proposed that capture-recapture techniques applied by population biologists to estimate species richness may deal successfully with song data (Catchpole and Slater 1995; Bell and Pledger 2002; Forstmeier and Balsby 2002; Garamszegi et al. 2002a; 2002b; Bell et al. 2004). Such approaches are based on an analogy with capture-recapture methods, because song or syllable types can be regarded as the units of capture-recapture. Song or syllable types have heterogeneous appearance in the repertoire, which is similar to the varying detection probability of species in an animal community or individuals in a population. However, the applicability of the proposed models in different species with different song organization has been little tested.

This paper describes a collaborative study to investigate the applicability of the suggested estimation procedure in 18 species with dissimilar song structure (Table 1). First, considering the structural organization of bird song, we established reliable standards for unit definition and sampling. Second, based on these schemes, we performed capture-recapture approaches to fit the estimation procedures with species-specific song organization, and estimated syllable repertoire size. Third, we compared our estimates with previously reported data that were obtained by simple enumeration applied in a less-standardized manner.

Sampling avian repertoires: units, sampling schemes and song organization

Sampling of bird song involves the definition of units and the selection of the songs on which the estimation will be based. These procedures require several steps that may render the prospective estimation sampling dependent. Hence, these potential problems should always be considered before performing any estimation. In Table 2, and in the following discussion, we summarize the most important, but frequently overlooked, problems that should be taken into account when defining units and sampling schemes for the estimation of song repertoires.

Estimation of repertoire size necessitates the definition of a biologically relevant unit (e.g. song types or syllable types as defined below) for which the expected abundance is desired. The chosen unit should represent a functional unit that is a target of selection, such as female choice, rather than an artificial fraction of the repertoire that is just easy to work with (Horn and Falls 1996; Horning et

Table 1 Summary statistics for the song data used, singing style (*Cont* continuous, *DiscontEV* discontinuous with eventual and *DiscontIMM* immediate variety) and previously reported individual syllable repertoire size in 18 bird species. Song recordings for this study were available from previous studies. We only used songs of unmanipulated males recorded in natural conditions. If the song of a particular species consists of two types of song (e.g. *Junco hyemalis*, *Hirundo rustica*, *Phylloscopus fuscatus*), types that are hy-

pothesized to function in mate attraction were considered. All recordings were subject to spectrographic analysis. References next to the species' name are for the conditions of song recordings. References next to different values are the sources of the particular data. A total of 2,429 syllable sequences (songs) originating from 171 individuals of 18 species were analyzed. On average we used 14.209 (SE=0.188) songs per male for each species

Species	No. individuals sampled	No. songs analyzed per individual (mean±SE)	Singing style	Previously reported syllable repertoire size
<i>Acrocephalus palustris</i> ¹	10	15±0	Cont	143.6 ¹
<i>Dendroica pensylvanica</i> ²⁻³	8	15±0	DiscontEV	16.1 ³
<i>Emberiza citrinella</i> ⁴	10	15±0	DiscontEV	4.1 ⁴
<i>Emberiza schoeniclus</i> ⁵	12	15±0	DiscontEV	30 ²⁹
<i>Ficedula albicollis</i> ⁶	10	15±0	DiscontIMM	31.4 ³⁰
<i>Ficedula hypoleuca</i> ⁷	10	15±0	DiscontIMM	27 ⁷
<i>Fringilla coelebs</i> ⁸⁻⁹	10	15±0	DiscontEV	not available
<i>Hirundo rustica</i> ¹⁰⁻¹²	16	7.8±0.9	DiscontEV	19.8 ¹⁰
<i>Junco hyemalis</i> ¹³⁻¹⁴	11	15±0	DiscontEV	2.4 ¹³⁻¹⁴
<i>Parus caeruleus</i> ¹⁵	10	15±0	DiscontEV	9.6 ³¹
<i>Parus major</i> ¹⁶	10	15±0	DiscontEV	7.9 ³²
<i>Phylloscopus fuscatus</i> ¹⁷⁻¹⁸	10	15±0	DiscontIMM	18.6 ¹⁷
<i>Phylloscopus trochilus</i> ¹⁹⁻²¹	10	15±0	DiscontIMM	29.7 ²¹
<i>Serinus canaria</i> ²²	8	15±0	DiscontIMM	69 ²²
<i>Sturnus vulgaris</i> ²³	6	15±0	DiscontIMM	39.9 ³³
<i>Sylvia communis</i> ²⁴⁻²⁵	11	15±0	DiscontIMM	616 ²⁵
<i>Turdus iliacus</i> ²⁶⁻²⁷	5	11±1	DiscontIMM	not available
<i>Turdus merula</i> ²⁸	4	15±0	DiscontIMM	not available

1) Bell et al. (2004); 2) Byers (1995); 3) Byers (1996); 4) Hansen (1984); 5) Nemeth (1996); 6) Garamszegi et al (2004); 7) Lampe and Espmark (1994); 8) Riebel and Slater (1999); 9) Riebel and Slater (2000); 10) Galeotti et al. (1997); 11) Galeotti et al. (2001); 12) Saino et al. (1997); 13) Titus (1998); 14) Titus et al. (1997); 15) Gorissen et al. (2002); 16) Van Duyse et al. (2002); 17) Forstmeier and Balsby (2002); 18) Forstmeier et al. (2002); 19) Gil et al. (2001) 20) Gil and Slater (2000a); 21) Gil and Slater (2000b); 22) Leitner et al. (2001); 23) Eens et al. (1991); 24) Balsby (2000); 25) Balsby and Dabelsteen (2001); 26) Espmark et al. (1989); 27) Lampe and Espmark (1987); 28) Draganoiu and L. Nagle, unpublished data; 29) E. Nemeth, unpublished PhD thesis; 30) Garamszegi et al. (2002a); 31) L. Gorissen, unpublished data; 32) E. Van Duyse, unpublished data; 33) Eens (1997)

Table 2 Potential issues to consider when defining sampling units and sampling schemes for the estimation of repertoire size. See text for details

Unit definition	Sampling
Biological relevance	Biological relevance
Reliability (inter observer repeatability)	Context
Biological question	Biological question
Song organization in terms of structural assemblage	Condition of recordings
	Selection of good quality recordings of songs
	Song organization in terms of singing style
	Consecutiveness versus randomness
	Number of songs analyzed

al. 1996; Searcy et al. 1999). If the unit of interest is wrongly defined, the estimation of "repertoire size" will not necessarily be a functionally important measure. Unfortunately, fundamental units remain to be identified in the majority of species (Searcy et al. 1999).

The unit of interest can be identified by visual inspection of sonagrams in which the units receive unique codes (Fig. 1). However, this classification involves some subjectivity, as it is based on certain criteria defined by the observer. Thus, the reliability of categorization should be tested by calculating inter-observer repeatability and reliability of scoring sonagrams by eye (Jones et al. 2001).

The appropriate unit should also be chosen from the recorded song according to the species and biological questions at hand, and further sampling schemes should be adjusted. Species-specific song organization will de-

termine the definition of the basic unit and the subsequent sampling protocol, thus finally the meaning of repertoire size. Bird song provides a natural arrangement of different sets of units that are grouped at different organization levels (Fig. 1).

The top-down organization of songs from the raw recording to the simplest elements may usually be apparent from observations, and the reliable unit should be selected on this basis (Fig. 1). Song performance may show structural patterns as determined by the temporal organization at particular time scales. To take a hypothetical example, the song performance may consist of some second-long songs separated by some second-long inter-song intervals, but more complex organization may also occur—e.g. if songs are arranged in bouts or choruses. Further, songs may contain some millisecond-long

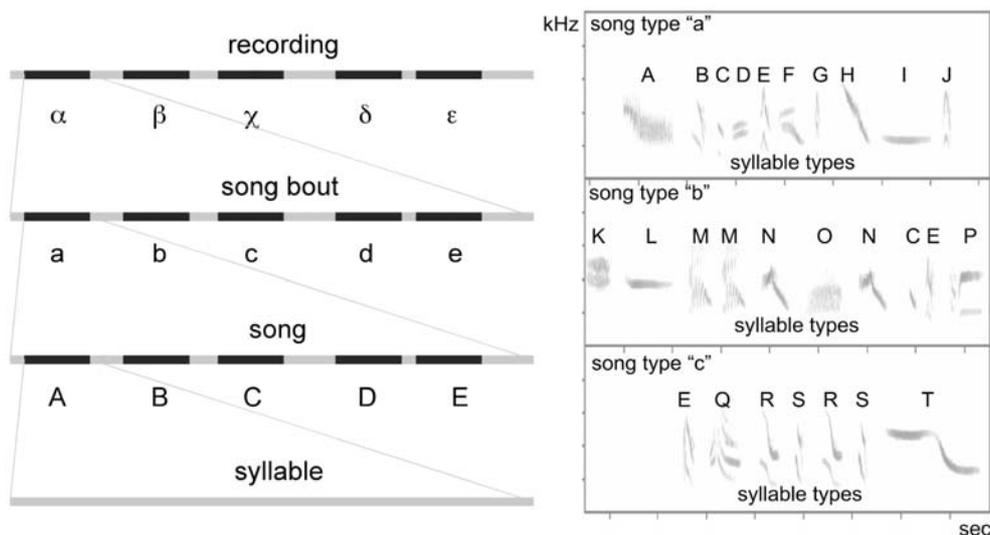


Fig. 1 Natural sampling schemes in the song of birds. Notes, figures, syllables, motifs etc. (A–E) as syllables are organized differently into higher level(s) of the song, which are the groups of elements separated by silent intervals. Phrases, strophes, sequences and songs can represent the next level of building blocks (a–e), but

subunits that are interrupted by some millisecond-long intervals. It is possible that on a smaller time-scale, these subunits can be divided into even smaller units of organization. This specific organization of songs provides us with natural sampling units that can all be the potential subjects of interest.

The subsequent repertoire size estimation will refer to the chosen sampling scheme, in which sampling units are the vocal units of interest, whereas sampling occasions are the higher organization levels along which the basic units are sampled. For example, song type repertoire size estimation corresponds to the number of unique song types (units) sampled across different recordings or song bouts (sampling occasions), while syllable repertoire size refers to the number of syllable types (units) across songs (sampling occasions). For simplicity, hereafter we will consider the ideal sampling as being for syllable repertoire size. We will use the term “repertoire size” for the number of unique syllable types in the whole recording or repertoire.

After defining units and sampling occasions, the next step is to select songs for further analysis that should be carefully adjusted for several factors (Table 2). Repertoire size may change with context, season, and age for instance. If all songs come from a specific period of the breeding season, then the resulting repertoire size estimate will pertain to this period, whereas songs collected over different behavioral states of the breeding and non-breeding seasons would yield inference about the bird’s entire repertoire. Similarly, the specific activity of the bird during recording of the songs is important. In some cases, the recordings may be standardized in the sense that they come from a bird engaged in a specific activity (e.g., sampling units might be a series of six consecutive songs), or they may be collected in various conditions and over various activities (e.g., three songs recorded in ter-

beyond this, further organization may occur (from bouts to complete dawn chorus (α – ϵ)). In this study we focus on syllables that are sampled across songs. Syllable types can be categorized by sonagram analysis, for which examples are given on the *right*

ritorial fights and four songs recorded during the dawn chorus). In addition, interest may focus not on an individual but on a group of individuals, for example of a particular population or species. In such cases, a song recorded from each of several individuals sampled from the group of interest would be an appropriate sampling unit. Other questions can be envisioned, and sampling schemes can be tailored to address them.

Songs for further analysis can be selected consecutively or randomly from a larger sample. Using consecutive songs may avoid heterogeneity caused by differences in the condition of recording. An additional advantage of this sampling is that some biological importance can be assigned to consecutive songs because, in nature, listeners are probably exposed to samples of consecutive songs. However, arguments about the biological relevance of a particular sampling method should be interpreted with care, because the sampling window of the intended receivers is unknown in most cases. The use of consecutive songs would at least yield more natural sampling than using random samples. Sometimes, the observer may prefer high quality songs with minimal background noise, which may help in the spectrogram analysis (e.g. Galeotti et al. 1997). However, this approach does not necessarily take all songs in their original order, so repertoire size estimation tends to involve more or less random sampling. When the purpose is to characterize differences among individuals in syllable repertoire size in repetitive singers, it may be uninformative to base comparisons of repertoires on consecutive songs only. Random samples, or sampling schemes based on other organization units, such as song types, may be useful in such cases.

The appropriate number of songs analyzed may vary from species to species. In general, the sample size should be met with the assumption that singers produce a vocal

versatility over the chosen sampling window that holds reliable information concerning the entire repertoire. Unfortunately, technological constraints, time and other limiting factors may often lead to a restricted number of available songs. The accurate sample size may be important, because in virtuoso species the more songs analyzed, the more syllable types one can detect, and thus any repertoire size estimation will be dependent on the number of songs sampled. This dependence may be especially important when relying on simple syllable counts. In this case, a particular aim of a study may be to compare individuals along a standard sampling design, for example based on 15 songs (this study) or 30-min recordings (e.g. Hiebert et al. 1989). Using reliable estimators for repertoire size, the effect of sample size on repertoire size may be less pronounced, however. Our study aims to achieve this by using capture-recapture approaches (see below).

In this study, we defined syllables on the spectrograms that are simple figures (or groups of figures) separated by noticeable time intervals (see also Fig. 1). With this definition, we intended to assign the smallest recognizable element in the sonagram that probably plays a functionally important role. As the biological relevance of syllables remains to be verified, we can at least infer that if syllable-like structures (so-called figures, notes or elements) can be recognized in many species, then they should have a biological function. Throughout the study we assumed that syllable categorization can be done with satisfactory reliability, which is probably the case in the majority of species as indicated by the high inter-observer repeatability of sonagram screening (e.g. Balsby and Dabelsteen 2001; Garamszegi et al. 2003).

We defined continuous vocalizations containing no longer than 1-s interruptions as songs. Hence, according to our definitions, syllables were the units of interest, which were sampled across songs, as sampling occasions. In each species, the proper application of these definitions was always verified by independent observers, and, if it was necessary, we redefined syllables or songs until reaching some agreed standard. Because the marsh warbler, *Acrocephalus palustris*, sings uninterruptedly, we could not define songs according to the above definition. For this species we adopted an alternative sampling approach that uses 6-s-long vocal units as sampling occasions of syllables (Bell et al. 2004).

To get comparable estimates, we attempted to apply the same sampling protocol for each species. Hence, the differences we expected to find should be due to differences in song organization and song complexity and not due to differences in sampling. Based on data availability, with the exception of two species (see Table 1), we used 15 consecutive songs per male. However, the assumption that this sampling window is representative of the whole repertoire may not be true in certain species or in certain contexts (see below). We are aware that more samples would be needed in some species for particular purposes. In repetitive singers, the use of 15 consecutive songs may not be sufficient to reflect full species-specific song or-

ganization. A small sample may contain only few syllable types with little variation, resulting in the underestimation of true syllable repertoire size independent of the estimation approach used.

Estimating avian repertoires from sequences of sampling occasions

Based on the established sampling design, we estimated syllable repertoire size by using capture-recapture approaches (Otis et al. 1978), based on models developed for closed populations. Because we investigated songs over a relatively short time window, we viewed repertoire size as fixed and the “population” of syllables as closed. However, some questions about bird song might involve longer time scales requiring open population models. For example, studies of song over time in young birds could be used to investigate learning, in which syllables are added to (or lost from) the repertoire. Robust design approaches could be used to deal with such open modeling (e.g. Pollock et al. 1990; Williams et al. 2002). We also note that other models that are able to handle presence-absence data and estimate the diversity of biological units can be envisioned for similar purposes. For instance Markov models that conduct comparisons of long repetitive sequences, such as nucleotide sequences (Rocha et al. 1999a, 1999b), might be useful for repertoire sequences. Any approach used for estimation will apply the assumption that the preceding unit definition and sampling was done in an unambiguous manner.

With capture-recapture models, the main objective is the estimation of abundance N , but appropriate modeling of p_{ij} , the probability of detection of the sampling unit i in sample j , is required to provide an accurate estimate of N (with low variance and low bias). Here, we are interested in estimating repertoire size (N) based on the appearance of syllables as sampling units (i) sampled across songs (j) as sampling occasions. Heterogeneous detection probability of syllables (p) refers to the phenomenon that some syllables are sung rarely and others are sung frequently. Otis et al. (1978) proposed a framework of eight models for p_{ij} , which differ in their assumptions regarding the sources of variation in detection probability (Table 3). Detection probability of syllables may be different for different syllables (called “heterogeneity”, h), but it may also vary between songs (called “time effect”, t), and/or it may depend on whether a particular syllable has or has not been sung previously (“behavior” effect, b). In the following paragraphs and in Table 3 we review the assumptions of these models as adapted for song repertoires.

The simplest model, $M(0)$, assumes p is constant, unaffected by heterogeneity, time or behavior (Otis et al. 1978). In the repertoire context, it assumes that different syllable types are equally likely to be sung in every song, and thus have the same probability of being detected on every sampling occasion. This estimator is not robust to any type of variability in detection probabilities; in particular there is negative bias when detection probability

Table 3 Notation, predictions and estimates of different capture-recapture models together with their implications for avian repertoire sizes. For simplicity, descriptions are based on a hypothetical sampling scheme, in which the units of interest are syllables that are sampled across songs as sampling occasions (see text for more details)

Model and the corresponding estimator available in CAPTURE	Predictions for syllable repertoires of birds	Source of variation in detection probability	Estimated parameters in CAPTURE
M(0) – “Null”	Simple songs with homogeneous structure	No variation	N : individual syllable repertoire size p : constant detection probability of syllables
M(t) – “Darroch”	Minor importance for songs in general, but may be sensitive to variable song length	Variation among songs	N : individual syllable repertoire size p_j : detection probability of syllables for each song
M(b) – “Zippin”	Specific repetitions and switching techniques may meet the assumptions (see text)	Variation due to pre-capture history of syllables	N : individual syllable repertoire size p : constant detection probability of undetected syllables c : constant detection probability of already-detected syllables
M(h) – “Jackknife”	High importance for songs, allows heterogeneous detection probability of syllables	Variation among syllables	N : individual syllable repertoire size p : averaged detection probability of syllables
M(tb) – “Burnham”	Minor importance for songs in general, the combination of predictions of models M(b) and M(t) is expected	Variation among songs and due to pre-capture history of syllables	N : individual syllable repertoire size p_j : detection probability of undetected syllables for each song $\theta = (\log_{p_j} c_j)^{-1}$ where c_j : detection probability of detected syllables for each song
M(bh) – “Removal”	Songs with specific repetitions (see model M(b)) and with variable detection probability of syllables; may be useful when syllable repertoire size depicts a cumulative curve with increasing sample size	Variation among syllables and due to their pre-capture history	N : individual syllable repertoire size No specific estimates of detection probability is available in CAPTURE
M(th) – “Chao”	May be sensitive to variable song length; specific sampling schemes (e.g. across individuals) may result in good data fit	Variation among songs and syllables	N : individual syllable repertoire size p_j : averaged detection probability of syllables for each song
M(tbh) – no estimator available in CAPTURE	Difficult to predict	Variation among songs and syllables and due to pre-capture history of syllables	No specific estimates of individual syllable repertoire size and detection probability is available in CAPTURE

varies among syllables. The assumptions of this model may be met in songs of species having simple songs with syllables of equal importance. However, birds that have both rare and common syllables, which is usually the case, may violate the assumption of homogeneity.

Model M(t) permits detection probability to vary between songs as sampling occasions, but all syllables have the same detection probability within songs (Darroch 1958; Otis et al. 1978). This model has a different capture probability p_j for each sample j . Hence, this model estimates repertoire size (N) and p_j detection probabilities for $j=1\dots K$ songs. It is difficult to regard this model as playing an important role in estimating repertoire size of bird song, because it would require songs with variable structure that result in detection probability varying across songs, but with the detection probability remaining homogenous among syllables. However, species with varying song length that inherently involves varying syllable detection probability may represent exceptions.

A similar model of somewhat restricted usefulness is M(b), which involves the possibility of behavioral responses of the units of interest (Otis et al. 1978; Colwell and Coddington 1994; Cam et al. 2002). This model arose in animal abundance studies, where all unmarked animals have the same probability of capture on any given trapping occasion, but all marked animals have another probability of capture. In other words, animals might move from a constant probability p for first capture to a new constant probability c for recapture; $c < p$ indicates trap shyness, and $c > p$ shows trap addiction. This model

does not allow for heterogeneity among animals or among samples. In a repertoire analysis, the three parameters in M(b) are N = repertoire size, p = probability of detecting a previously unobserved syllable in a song (“capturing an unmarked animal”), and c = probability of detecting a previously-detected syllable. There are at least two plausible effects that may cause some songs to require a model including a behavior effect [see also models M(tb), M(bh), M(tbh) below]. The first is associated with the observer’s influence on analyzing sonagrams, because the ability to identify a syllable as a new or existing category may be determined by previous experience. This will be a problem only in species with large repertoires when the catalogue system becomes complicated. The second, biologically more important source of heterogeneity due to behavioral effects may be caused by repetitions and specific song switching styles. When a bird switches to a new syllable type that is repeated several times before switching to another, detection probability will increase for this new syllable.

Heterogeneity in detection probabilities of syllables is a general phenomenon in birds, because apparently certain syllables are more commonly used than others (Garamszegi et al. 2002a; Bell et al. 2004). These conditions meet the assumption of the model M(h) of Otis et al. (1978) that permits such heterogeneity of detection probabilities among syllables, but assumes that there is no other variation, e.g. among the different songs. We demonstrated the use of this model by using the associated jackknife estimator of Burnham and Overton (1978,

1979) to estimate N in the collared flycatcher, *Ficedula albicollis*, and in the marsh warbler, (Garamszegi et al. 2002a; Bell et al. 2004).

The other four models of Otis et al. (1978), $M(tb)$, $M(th)$, $M(bh)$ and $M(tbh)$ contain combinations of the t , b and h effects. These are used if two or more of the effects occur simultaneously. When time and behavioral patterns are present in the data, model $M(tb)$ may prove useful (Otis et al. 1978). This model has two sequences of capture probabilities over successive songs, p_j for a first detection and c_j for a redetection ($j=1\dots K$ songs). c_j is a function of initial capture probability, with the relationship given by $c_j=p_j^{1/\theta}$. Parameters estimated are repertoire size, N , initial detection probabilities p_i and θ . This model would be used if there were a combination of specific switching techniques of syllables together with different song structures requiring a time effect. However, this model also assumes all syllables follow the same detection probabilities, a sequence of p_1, p_2 , etc. for the early samples up until first capture, followed by a switch to the c_j series. Since the model fails to allow for heterogeneity among syllables (h), it is unlikely to be of general importance.

In general, the assumption of equal detection probability among syllables in models $M(0)$, $M(t)$, $M(b)$ and $M(tb)$ will presumably restrict their applicability for song complexity. This assumption is relaxed in models that involve the heterogeneous appearance of syllables in the repertoire. Model $M(th)$ is the approximate probability model when the variation in detection probability is associated with individual syllables and also with different samplings, such as songs (Otis et al. 1978; Chao et al. 1992). This model can estimate the repertoire size, N , and the average detection probability in each song p_j . Model $M(th)$ may prove useful where different songs have specific functions, for example one type is for territory defense and another for mate attraction, and we use the mixture of these songs for repertoire size estimation. In this case, it might be expected that some syllables are sung more frequently in one context, but rarely in the other one. Changeable song length may result in heterogeneous syllable detection probability between songs.

Model $M(bh)$ assumes that detection probability is subject to individual variation and to behavioral response (Otis et al. 1978; Pollock and Otto 1983; Norris and Pollock 1995, 1996). Parameters estimated are the repertoire size and the average detection probabilities for previously unobserved syllables in each song. These averages are expected to vary among songs because previously unobserved syllables have declining detection probabilities as the number of occasions increases (the syllables with high detection probabilities are observed early in the sampling). Specific repetition patterns in the data may lead to samples fitting the assumptions of model $M(bh)$. This model may be used in cases where available data are in the form of syllable accumulation curves depicting the number of new syllables, as the sample size increases. The estimators for this model permit estimation for this abbreviated form of syllable data and have been

recommended for use with the analogous problem of species richness estimation based on the accumulation of previously unrecorded species (Nichols and Conroy 1996; Cam et al. 2002). Ideally when estimating repertoire size, a plot of the number of new syllables found against the number of songs analyzed forms an accumulation curve (Wildenthal 1965; Catchpole and Slater 1995). The accumulation curve approaches an asymptote at N , where only the less common syllables are being newly found, as the more common elements will have been found earlier. In species with large repertoires it is predictable that by analyzing 15–50 songs only, one can be still far away the asymptote, i.e. from the real number of syllables known by the individual. This kind of situation will require use of estimators based on model $M(bh)$ rather than visual inspections of accumulation curves.

For conceptual purposes it is important to mention the model $M(tbh)$, which allows detection probabilities to vary by behavioral response, sampling occasion and individual syllable. This estimator is not available in the most frequently used statistical software (CAPTURE, Rexstad and Burnham 1991) that we used as well, thus no parameter estimates are available under this model. Including all three factors that may affect detection probabilities in one model precludes the model from having any practical use in the estimation of repertoire size. If the experimenter is led to regard the model $M(tbh)$ to be required for a given data set, Otis et al. (1978) suggest reevaluation of the experiment to obtain useful statistical information concerning the parameters of interest that can be estimated by previous models. For songs it would mean the revision and modification of the sampling scheme. New estimators have been developed for this model but have seen little use (Lee and Chao 1994; Pledger 2000).

The availability of diverse models requires an objective approach to model selection for the data set at hand. For this purpose Otis et al. (1978) describe an algorithm using a discriminant function analysis built with simulated data, which is based on test statistics from between-model tests and goodness-of-fit tests. After selecting the most appropriate model, it is possible to obtain estimates for this model.

Model selection, detection probability and song organization

We analyzed individual syllable sequences by the software CAPTURE (Rexstad and Burnham 1991) that offers a model selection algorithm for the eight closed population models (Otis et al. 1978) and under various models, computes estimates of abundance and detection probabilities of sampling units together with the associated standard errors. The results of the model selection and repertoire size estimations for all species involved are summarized in Table 4. Our aim in this study was to investigate how capture-recapture approaches deal with song data in different species with varying song organi-

Table 4 Means (\pm SE) of enumerations and estimations of individual repertoire sizes and detection probability of syllables based on 15 consecutive songs in 18 passerine species. Estimated and enumerated repertoire sizes were compared with previously reported values (see Table 1). Estimated repertoire size is from the model offering the highest data fit in each male. Syllable repertoire size estimations could not be obtained for individuals for which model M(tbh) has been selected. The mean of the estimated syllable

repertoire size was calculated by using individuals only, for which estimations were available based on another best model. In the chaffinch *Fringilla coelebs*, model M(tbh) has been selected globally, the estimations in the table rely on individuals, for which model M(h) was the best model. Due to different sample sizes, estimated repertoire size may be smaller than the enumerated. When more than one model is given, these models have been selected with equal frequency within the species

Species	Enumerated syllable repertoire size	Model selected	Estimated syllable detection probability	Estimated Syllable repertoire size	Reported vs. Enumerated	Reported vs. Estimated
<i>Acrocephalus palustris</i>	113.6 \pm 6.8	M(t)	0.09 \pm 0.04	241.1 \pm 45.0	$t_9=-4.4$ $p<0.01$	$t_9=2.2$ $p=0.06$
<i>Dendroica pensylvanica</i> ^a	5.5 \pm 0.7	M(0)	1.00 \pm 0.00	5.4 \pm 0.8	$t_7=-16.2$ $p<0.01$	$t_6=-14.2$ $p<0.01$
<i>Emberiza citrinella</i> ^a	4.4 \pm 0.2	M(0)	0.57 \pm 0.05	4.5 \pm 0.2	$t_9=1.8$ $p=0.10$	$t_7=2.1$ $p=0.07$
<i>Emberiza schoeniclus</i> ^a	6.6 \pm 0.5	M(bh)	0.84 \pm 0.09	6.3 \pm 0.6	$t_{11}=-43.1$ $p<0.01$	$t_9=-38.50$ $p<0.01$
<i>Ficedula albicollis</i>	30.4 \pm 2.5	M(h)	0.17 \pm 0.02	42.2 \pm 6.5	$t_9=-0.4$ $p=0.69$	$t_8=1.6$ $p=0.14$
<i>Ficedula hypoleuca</i>	25.2 \pm 3.4	M(h)	0.22 \pm 0.05	29.8 \pm 6.5	$t_9=-0.5$ $p=0.61$	$t_7=0.4$ $p=0.69$
<i>Fringilla coelebs</i>	19.4 \pm 2.0	M(tbh) (M(h))	0.53 \pm 0.10	17.8 \pm 3.6		
<i>Hirundo rustica</i>	19.7 \pm 0.4	M(0)	0.77 \pm 0.04	20.2 \pm 0.6	$t_{15}=-0.3$ $p=0.75$	$t_{11}=0.7$ $p=0.52$
<i>Junco hyemalis</i> ^a	1.2 \pm 0.1	M(0)	1.00 \pm 0.00	1.2 \pm 0.1	$t_{10}=-9.6$ $p<0.01$	$t_{10}=-9.6$ $p<0.01$
<i>Parus caeruleus</i> ^a	3.4 \pm 0.4	M(0)	0.89 \pm 0.07	3.4 \pm 0.4	$t_9=-15.5$ $p<0.01$	$t_9=-15.5$ $p<0.01$
<i>Parus major</i> ^a	3.4 \pm 0.4	M(0)	0.79 \pm 0.09	3.4 \pm 0.4	$t_9=-10.2$ $p<0.01$	$t_9=-10.2$ $p<0.01$
<i>Phylloscopus fuscatus</i>	12.0 \pm 0.8	M(0)	0.09 \pm 0.02	25.3 \pm 9.3	$t_9=-8.4$ $p<0.01$	$t_9=0.72$ $p=0.49$
<i>Phylloscopus trochilus</i>	28.1 \pm 1.3	M(h)	0.24 \pm 0.02	35.1 \pm 2.9	$t_9=-1.2$ $p=0.27$	$t_7=1.9$ $p=0.10$
<i>Serinus canaria</i>	63.3 \pm 2.0	M(tb)	0.23 \pm 0.03	70.7 \pm 3.8	$t_7=-2.2$ $p=0.02$	$t_7=0.4$ $p=0.68$
<i>Sturnus vulgaris</i>	41.0 \pm 5.7	M(t) M(th)	0.29 \pm 0.03	43.7 \pm 6.7	$t_5=0.2$ $p=0.86$	$t_5=0.6$ $p=0.60$
<i>Sylvia communis</i>	58.7 \pm 6.8	M(h)	0.11 \pm 0.02	118.5 \pm 21.8	$t_{10}=-81.4$ $p<0.01$	$t_9=-22.8$ $p<0.01$
<i>Turdus iliacus</i>	100.6 \pm 28.0	M(h) M(t)	0.08 \pm 0.04	685.2 \pm 501.8		
<i>Turdus merula</i>	57.5 \pm 11.9	M(h) M(0)	0.07 \pm 0.02	107.8 \pm 26.2		

^a Species with small syllable repertoire size in which capture-recapture estimates should be used with caution (see text for details)

zation and potential repertoire size, thus we performed model selection and estimation procedures for each species involved. We note, however, that capture-recapture models were developed for large population sizes (Otis et al. 1978; Nichols and Conroy 1996), i.e. repertoire sizes in our context, thus any computation in species with small repertoire size should be handled with caution and make the estimation of repertoire size and detection probabilities conservative. Means and standard errors are given below in the main text and in the figures, whereas in the statistical analyses syllable repertoire sizes were \log_{10} -transformed, but detection probability was arcsine-square-root transformed.

Model selection may be a function of sampling but can also be influenced by biological characteristics. Since birds with elaborate songs usually have common and rare syllables in their repertoire, we predicted that syllable heterogeneity should be the most common source of

variation in detection probability, thus models assuming this variability (M(h), M(th), M(bh), M(tbh)) should be the most useful models for estimating syllable repertoire size and related parameters. In our sample of eighteen species, no model was selected clearly more often than any other of the seven alternative models. The most frequent models selected were M(0) and M(h) (selection probabilities: 35% and 25% respectively). Models including heterogeneity in the detection probability among individual syllables (M(h); M(bh); M(th); M(tbh)) accounted for 54% of the selected models. The overall proportion of goodness-of-fit tests of model M(h) with P -values >0.05 was 71%. The proportion of individual series of syllable sequences for which M(h) was selected varied among species ($\chi^2=68.9$, $df=17$, $P<0.01$).

We expected that model selection would be a species-specific attribute, because it should be sensitive to specific song organization patterns. Therefore, we predicted

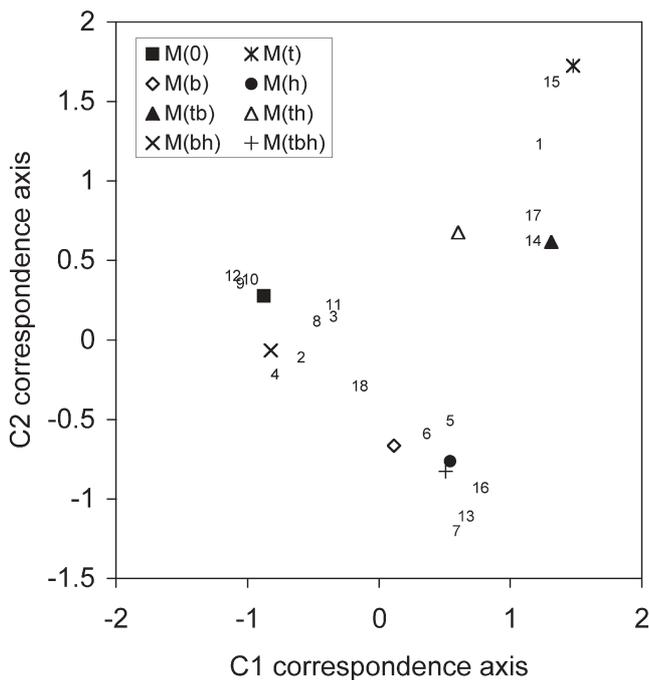


Fig. 2 Correspondence analysis of eighteen bird species along the capture-recapture model selection results based on their song structure. Numbers are for species: 1: *Acrocephalus palustris*; 2: *Dendroica pensylvanica*; 3: *Emberiza citrinella*; 4: *Emberiza schoeniclus*; 5: *Ficedula albicollis*; 6: *Ficedula hypoleuca*; 7: *Fringilla coelebs*; 8: *Hirundo rustica*; 9: *Junco hyemalis*; 10: *Parus caeruleus*; 11: *Parus major*; 12: *Phylloscopus fuscatus*; 13: *Phylloscopus trochilus*; 14: *Serinus canaria*; 15: *Sturnus vulgaris*; 16: *Sylvia communis*; 17: *Turdus iliacus*; 18: *Turdus merula*. See text for explanations

that in different species different capture-recapture models will offer the best fit with the data depending on song organization. In accordance with this prediction, we found that the observed selection probabilities differed significantly from expectations based on random choice ($\chi^2=282.1$, $df=119,164$, $P<0.01$).

We arranged species based on the associated model selection results for individuals along the eight CAPTURE models in a two-dimensional space by applying a correspondence analysis (Fig. 2). In general, the correspondence analysis describes the relationships between two nominal variables in a correspondence table in a low-dimensional space, where the distances between category points in a plot reflect the relationships between categories of the same or of different variables (Greenacre 1993). In particular, the correspondence analysis represents species and CAPTURE models as occurring in a postulated ordination space, where the distance between a particular species and a particular model reflects how the given model fits the song structure of the given species. This analysis revealed that species could be well separated based on the model selection results (Fig. 2).

To seek a relationship between model selection and species-specific song organization, we tested whether model selection results are associated with the major singing styles that have been described in birds, and are

related to the temporal arrangement of songs (Hartshorne 1956, 1973; Kroodsma 1982). Continuous (i.e. warbling) singers emit long vocal sequences without interruptions, so that the raw recording may merely consist of syllables without being split into songs (ABCDEFGHIJKLMNOPQR...). On the other hand, discontinuous singers present their songs with silent inter-song intervals. Discontinuous singers may sing with immediate or eventual variety. Birds singing with eventual variety repeat the same within-song vocal pattern before switching to another song type (eventual variety with repetitive syllables: AAA AAA AAA BBB BBB BBB; eventual variety with multiple syllables/song: ABC ABC ABC DEF DEF DEF). Birds of species that sing a vocal sequence once and then immediately switch to the next are known as singers of immediate variety (immediate variety with repetitive syllables: AAA BBB CCC DDD EEE FFF; immediate variety with multiple syllables/song: ABC DEF HIJ KLM MNO PQR). We predicted that for birds having similar song organization, model selection statistics will give similar results.

The separation of species according to the model selection results, as revealed by their position in Fig. 2, may represent species-specific patterns of song organization. Relying on singing style categories of birds (Table 1), we found significant differences in C1 correspondence axis among birds singing continuously, discontinuously and with eventual variety, and discontinuously and with immediate variety (Kruskal-Wallis ANOVA: C1 axis, $\chi^2=7.7$, $df=2$, $P=0.02$, Fig. 3a; C2 axis, $\chi^2=2.1$, $df=2$, $P=0.35$). A comparison of species for which models M(0) and M(h) had been selected showed significant differences in their singing styles ($\chi^2=9.4$, $df=1,10$, $P<0.01$; the European blackbird, *Turdus merula*, was excluded from this test, because in our small sample, both models M(0) and M(h) have been selected with equally high probability, see Table 4 and Fig. 2). Hence, for birds singing with eventual variety model M(0) has been selected with high probability for our samples. For singers of immediate variety model M(h) and models including “time” effects (M(tb), M(th), M(t)) performed better. Additionally, model selection results showed some intraspecific variation among individuals as well. Hence, when combined with adequate sampling, the model selection algorithm may be sensitive to individual- or species-specific song organization patterns.

For each individual we obtained estimates for syllable repertoire size and detection probability by using the model that, based on the selection routine, offers the best fit to its song. Under the best model, average detection probabilities of syllables ranged from 0.009 to 1.000 (Mean \pm SE=0.45 \pm 0.03). Detection probability was significantly repeatable across individuals within species ($F_{17,133}=48.6$, $P<0.01$, $R=0.87$, based on Lessells and Boag 1987). Therefore we calculated mean syllable detection probabilities for each species assuming that this is a species-specific attribute. This detection probability was also related to the mode of singing, as species singing with eventual variety had larger average detection prob-

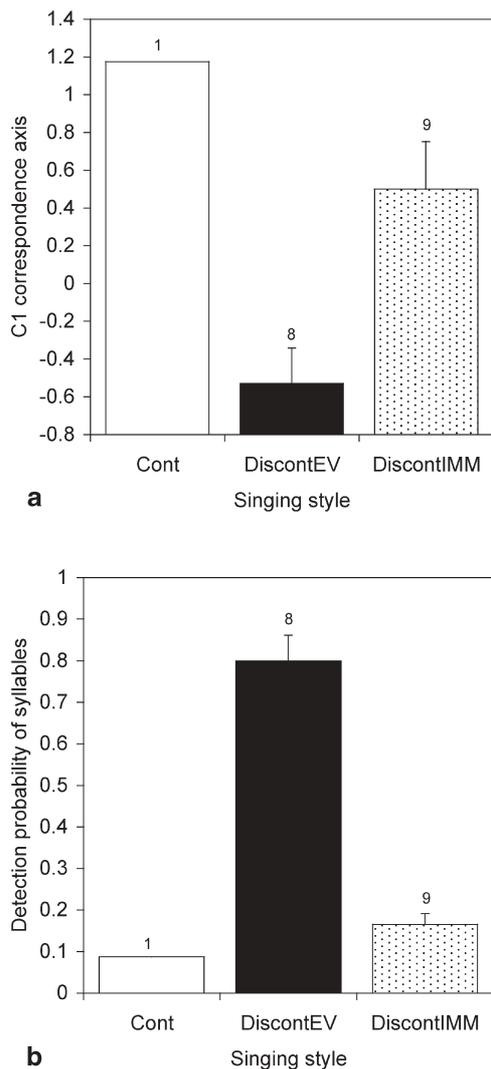


Fig. 3 **a** Model selection results and singing styles of birds. **b** Detection probability of syllables and singing styles in birds. Singing styles: *Cont* continuous singers, *DiscEV* discontinuous singers with eventual variety, *DiscIMM* discontinuous singers with immediate variety (see Table 1). C1 correspondence axis is from Fig. 2 representing the outcome of model selection. Sample sizes (*numbers*) and standard errors (*bars*) are given

ability for their syllables than species singing with immediate variety (Kruskal-Wallis ANOVA: $\chi^2=12.9$, $df=2$, $P<0.01$; Fig. 3b).

Species with small repertoires probably have equal detection probabilities for their syllables, which have the same chance of being sung in every song. Small repertoire size thus leads to simpler estimators, and for these birds the model M(0) was frequently selected (Table 4). Homogeneous detection probability of syllables may be observed if (1) males sing a large proportion of their entire repertoire in a single song (e.g. chestnut-sided warbler *Dendroica pensylvanica*, barn swallow *Hirundo rustica*); or (2) males repeat the same fixed simple sequence for a couple of songs then they switch to another sequence (e.g. yellowhammer *Emberiza citrinella*, great

tit *Parus major*, blue tit *P. caeruleus*). Dark-eyed junco (*Junco hyemalis*) males usually have a single syllable, thus the song performance may be regarded as the combination of both. The dusky warbler *Phylloscopus fusca-tus* has a relatively large repertoire, and birds tend to “repeat” one syllable type for one song and immediately switch to another for the next song resulting in the low average detection probability of syllables. Within the M(0) species, birds singing a large amount of the repertoire in a particular song are expected to have higher detection probabilities (e.g. chestnut-sided warblers: 1.00, barn swallows: 0.77) than repetitive singers (e.g. yellowhammer: 0.57, dusky warbler: 0.09). Note that by using 15 consecutive songs for the tit species, song type switches were often missed, thus the resulting samples had the same syllable sequences causing relatively high detection probability (great tit: 0.89, blue tit: 0.79). In addition to the inappropriate sampling, some caution is needed when interpreting these results, because it has been suggested that capture-recapture approaches do not perform well when species abundance (or syllable repertoire size in our case) is small (Otis et al. 1978).

The collared and pied flycatcher *Ficedula hypoleuca*, the willow warbler *Phylloscopus trochilus* and the white-throat *Sylvia communis* were found to be typical M(h) species [for the European blackbird only four individuals were analyzed for which both models M(0) and M(h) showed high selection statistics]. These species tended to have larger syllable repertoires, and apparently they organize their songs with heterogeneous probabilities of syllable use. There appear to be some syllable types that are more frequently sung than others. These syllables may have different functions. Common syllables might be used in fixed sequences, such as introductory or terminal motifs, while rare syllables might result, for example, from improvisation. However, probably the most important source of heterogeneity in detection probability of syllables is that the cost associated with the production of different syllables may vary. Costly syllables are expected to become reliable signals of male quality, thus favored by sexual selection.

In the reed bunting *Emberiza schoeniclus*, which has a repetitive song performance, model M(bh) was the most frequently selected model, suggesting that a “behavioral” -effect may be prominently present in its song. Using model M(b), which also provided a relatively good fit to the data, the detection probability p of previously unobserved syllables was lower than the detection probability c of previously detected syllables. This result suggests a “trap-happy” behavioral response, in which once a syllable type is sung in a song the probability of it being sung in the next song is higher. In the reed bunting fixed introductory sequences can be found at the beginning of the song, while the end is rather variable (Nemeth 1996). The heterogeneous detection probability of syllables, in combination with the specific repetition of the terminal elements over a couple of songs, may be why it best fits model M(bh).

For some species, members of the “t”-model family (M(t), M(tb), M(th)) tended to be selected, implying that the detection probability of syllables is different in different songs (marsh warbler, canary *Serinus canaria*, European starling *Sturnus vulgaris*, redwing *Turdus iliacus*). This pattern might occur because song performance is variable. The inspection of syllable sequences allows us to conclude in these species that song length in terms of the number of syllables sung is changeable, which might cause the detection probability to vary across songs. In longer songs syllables are likely to have a higher probability of being sung. The estimation results for the marsh warbler should be interpreted with care, because we used artificial syllable sequences for 6-s-long “songs” (when we used other criteria for song length, the model selection and the subsequent estimation were quite different for the same sequences).

For male chaffinches *Fringilla coelebs*, the model M(bh) was selected with the highest probability, but model M(h) also showed a relatively high selection statistic. In Fig. 2 this species was grouped with M(h) species. The songs of the chaffinch consist of repeated sequences of some syllables followed by a terminal motif, and this species has fixed sequence of syllable types within a particular song type (Slater 1981). The specific outcome of the model selection may result from the varying repetition of particular song and syllable types, which may raise heterogeneity, time and behavioral effects together.

At this moment, we should be careful with interpreting how different closed population models explain vocal structures, because capture-recapture models have been originally developed to deal with species lists of surveys of animal communities. Specific patterns present in the data may be different for songs and animal communities. For example a special problem present in songs is the non-independence of syllables (Bell et al. 2004). Capture-recapture methods assume that the appearances of the units of interest across sampling occasions are independent of each other. This assumption will be violated in cases in which syllables are associated, resulting in non-random sequences of syllables in the song. For example, it may happen that one syllable type frequently follows another leading to their non-independent appearance. In this case, one should consider statistical ways of estimating variances of repertoire estimates caused by such non-independence. One approach would involve use of quasi-likelihood procedures for overdispersion to compute variance inflation factors based on model goodness-of-fit tests (see Burnham et al. 1987; Lebreton et al. 1992). Another approach would be to estimate variances using a bootstrap approach (e.g. Nichols et al. 1998). Such variance estimation in the face of non-independence merits additional study. Based on our results, however, we believe that the further development of classical closed population models toward dealing with specific issues associated with songs offers fruitful solutions in handling species-specific song attributes.

Counting of syllables versus capture-recapture approaches

In most studies of avian song repertoires, count data are used as an estimate of repertoire size, which disregards sources of variation in the use of syllables and assumes that the detectability of different syllables is equal. However, we showed that heterogeneous detection probability of syllables was present in most individuals and species, and varied among species depending on the song organization. Studies relying on syllable enumeration do not detect all the syllables in a given song repertoire in a relatively small sample, and cannot evaluate the missed fraction of repertoire size. The probability that each syllable will be sung and detected may vary among individuals, populations and species, but also among syllables and conditions of recording. Hence, different probabilities of detection among syllables can generate a number of negative biases of unknown magnitude when repertoire size is computed via naïve counting of syllables. The unknown missed fraction of repertoire size may give false results in comparisons of repertoire size if the missed fraction differs between the units of comparison, such as individuals, populations, and species.

We compared our estimates taking differences in detection probability into account with previously reported data on syllable repertoire size that were obtained by heterogeneous sampling approaches and classical enumeration procedures based on much larger samples (see data in Table 1). The estimated syllable repertoire size based on the best models ranged from 1 to 2684 (mean±SE: 65.3±19.1). The observed syllable repertoire size in the same sample varied from 1 to 199 (mean±SE: 29.1±2.6), whereas previously reported individual syllable repertoire size ranged from 2.35 to 390 (Table 1). In our samples, syllable repertoire size was significantly repeatable across individuals within species indicating that it showed consistent variation within species and it can be a species-specific attribute (based on simple enumeration: $F_{17,170}=155.8$, $P<0.01$, $R=0.94$; using capture-recapture estimates: $F_{17,146}=66.7$, $P<0.01$, $R=0.89$).

A simple enumeration of syllable types in 15 consecutive songs is likely to underestimate true syllable number as determined by previous observations using larger samples. Relying on the individual enumerations, we applied t -tests to check whether the mean enumerated repertoire size in different species significantly differed from the previously reported mean value (Table 4). The signs for the corresponding t -tests were negative in thirteen out of fifteen cases ($\chi^2=9.0$, $df=1$, $P<0.01$) indicating that enumerations based on 15 songs were underestimations in relation to previous syllable counts from larger samples. Note that where t -tests were significant, they always showed that enumerations in our small sample were generally lower than previously reported.

However, estimates of individual syllable repertoire size given by the model best fitted to the data tended to be higher than results of previous reports. In nine out of fifteen comparisons the sign of the associated t -tests were

positive ($\chi^2=0.6$, $df=1$, $P=0.44$). We also conducted paired t -tests in which we compared previously reported estimates with our estimates across species. This approach indicated that a simple enumeration gives smaller values for syllable repertoire size than previous observations drawn from larger samples ($t_{14}=-3.1$, $P<0.01$), whereas capture-recapture procedures did not give statistically different results from earlier reports ($t_{14}=1.9$, $P=0.08$). Hence, capture-recapture estimations of repertoire size from small samples have the potential to make more realistic estimations from small samples. When capture-recapture estimates indicate larger values for syllable repertoire size than previous reports, we can expect that by analyzing more songs, previous studies would have had a high chance of finding new syllable types in a larger sample. On the other hand, in species with small repertoire size and homogenous detection probability, the estimated syllable repertoire size does not differ from the outcome of a simple syllable count from the same sample (Table 4). This finding may be due to the inaccurate sampling and the limited usefulness of capture-recapture approaches for small repertoire. Nevertheless, if there are very few syllable types only, an observer may have a very high chance to find all of these in a small sample of song independent of the method used.

In conclusion, capture-recapture approaches may be superior to the simple enumeration of syllables, because they can handle patterns of song structure e.g. in terms of heterogeneity in detection probability. Therefore, capture-recapture approaches may provide better estimates than simple enumeration for syllable repertoire size in species with large repertoires and with heterogeneous syllable detection probability even from small samples. In species with small repertoire size and homogenous syllable use, counting syllables may be satisfactory, but only when combined with unambiguous sampling.

Selection forces favoring the evolution of song repertoires may also affect other song traits and vice versa. This is because an increase in individual repertoire size may also lead to a change in song organization resulting in a link between what and how birds sing. For example, larger repertoires can be arranged in longer songs or in songs with higher versatility (e.g. Eens 1997). The arrangement of syllables within songs will thus be altered and so influence their average detection probability. As a result, the evolution of syllable repertoire size (i.e. song complexity) and syllable detection probability (i.e. song organization) may be expected to occur in concert. Considering the listeners' perspective in vocally virtuoso species, it may be difficult for females to assess the entire repertoire size of the singers based on small samples. Therefore, it is plausible that listeners to complex songs do not count repertoire size, but assess syllable detection probability from small samples, in a similar way to capture-recapture approaches. Little is known about how repertoires function biologically during the perception of song. Researchers use repertoire size as a measure of song complexity, but maybe listeners adjust their response to songs by using other cues, such as syllable detection probability, that

correlates with syllable repertoire size. Capture-recapture approaches are able to deal with both quantities.

Acknowledgements The authors are highly indebted to all the colleagues, students and assistants who participated in the song recordings and analyses. We are very grateful to J. Nichols, S. Pledger, K. Riebel and P.J.B. Slater for their useful comments. K. Riebel kindly provided data on the chaffinch.

References

- Balsby TJS (2000) Song activity and variability in relation to male quality and female choice in whitethroats *Sylvia communis*. *J Avian Biol* 31:56–62
- Balsby TJS, Dabelsteen T (2001) The meaning of song repertoire size and song length to male whitethroats *Sylvia communis*. *Behav Proc* 56:75–84
- Bell BD, Pledger S (2002) Estimating the size of the complex song repertoire of the European marsh warbler *Acrocephalus palustris* using capture-recapture methods. *Proc Int Ornithol Congr* 23:165
- Bell BD, Borowiec M, Lontkowski J, Pledger S (2004) Short records of marsh warbler (*Acrocephalus palustris*) song provide indices that correlate with nesting success. *J Ornithol* 145:8–15
- Burnham KP, Overton WS (1978) Estimation of the size of closed population when capture probabilities vary among animals. *Biometrika* 65:625–633
- Burnham KP, Overton WS (1979) Robust estimation of population size when capture probabilities vary among animals. *Ecology* 60:927–936
- Burnham KP, Anderson DR, White GC, Brownie C, Pollock KH (1987) Design and analysis methods for fish survival experiments based on release-recapture. *Am Fish Soc Monogr* 5:1–437
- Byers BE (1995) Song types, repertoires and song variability in a population of chestnut-sided warblers. *Condor* 97:390–401
- Byers BE (1996) Geographic variation of song form within and among chestnut-sided warbler populations. *Auk* 113:288–299
- Cam E, Nichols JD, Sauer JR, Hines JE (2002) On the estimation of species richness based on the accumulation of previously unrecorded species. *Ecography* 25:102–108
- Catchpole CK, Slater PJB (1995) Bird song: biological themes and variations. Cambridge University Press, Cambridge
- Chao A, Lee SM, Jeng SL (1992) Estimation of population size for capture-recapture data when capture probabilities vary by time and individual animal. *Biometrics* 48:201–216
- Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. *Philos Trans R Soc Lond B* 345:101–118
- Darroch JN (1958) The multiple-recapture census. I. Estimation of a closed population. *Biometrika* 45:343–359
- Eens M (1997) Understanding the complex song of the European starling: an integrated ethological approach. *Adv Study Behav* 26:355–434
- Eens M, Pinxten R, Verheyen RF (1991) Male song as a cue for mate choice in the European starling. *Behaviour* 116:210–238
- Espmark YO, Lampe HM, Bjerke TK (1989) Song conformity and continuity in song dialects of redwings *Turdus iliacus* and some ecological correlates. *Ornis Scand* 20:1–12
- Forstmeier W, Balsby TJS (2002) Why mated dusky warblers sing so much: territory guarding and male quality announcement. *Behaviour* 139:89–111
- Forstmeier W, Kempenaers B, Meyer A, Leisler B (2002) A novel song parameter correlates with extra-pair paternity and reflects male longevity. *Proc R Soc Lond B* 269:1479–1485
- Galeotti P, Saino N, Sacchi R, Møller AP (1997) Song correlates with social context, testosterone and body condition in male barn swallows. *Anim Behav* 53:687–700
- Galeotti P, Saino N, Perani E, Sacchi R, Møller AP (2001) Age-related song variation in male barn swallows. *It J Zool* 68:305–310

- Garamszegi LZ, Boulinier T, Møller AP, Török J, Michl G, Nichols JD (2002a) The estimation of size and change in composition of avian song repertoires. *Anim Behav* 63:623–630
- Garamszegi LZ, Doutrelant C, Lampe HM, Boulinier T (2002b) ISBE round table report: a new method to estimate repertoire size. *ISBE Newsletter* 14:23–24
- Garamszegi LZ, Møller AP, Erritzøe J (2003) The evolution of immune defense and song complexity in birds. *Evolution* 57:905–912
- Garamszegi LZ, Møller AP, Török J, Michl G, Péczely P, Richard M (2004) Immune challenge mediates vocal communication in a passerine bird: an experiment. *Behav Ecol* 15:148–157
- Gil D, Gahr M (2002) The honesty of bird song: multiple constraints for multiple traits. *Trends Ecol Evol* 17:133–141
- Gil D, Slater PJB (2000a) Multiple song repertoire characteristics in the willow warbler (*Phylloscopus trochilus*) correlations with female choice and offspring viability. *Behav Ecol Sociobiol* 47:319–326
- Gil D, Slater PJB (2000b) Song organisation and singing patterns of the willow warbler, *Phylloscopus trochilus*. *Behaviour* 137:759–782
- Gil D, Cobb JLS, Slater PJB (2001) Song characteristics are age dependent in the willow warbler, *Phylloscopus trochilus*. *Anim Behav* 62:689–694
- Gorissen L, Janssens E, Pinxten R, Eens M (2002) Differences in song repertoire size and composition between two populations of blue tits *Parus caeruleus*. *Avian Sci* 2:39–47
- Greenacre MJ (1993) Correspondence analysis in practice. Academic Press, London
- Hansen P (1984) Neighbor-stranger song discrimination in territorial yellowhammer *Emberiza citrinella* males, and comparison with responses to own and alien song dialects. *Ornis Scand* 15:240–247
- Hartshorne C (1956) The monotony-threshold in singing birds. *Auk* 83:176–192
- Hartshorne C (1973) Born to sing. an interpretation and world survey of bird song. Indiana University Press, Bloomington
- Hiebert SM, Stoddard PK, Arcese P (1989) Repertoire size, territory acquisition and reproductive success in the song sparrow. *Anim Behav* 37:266–273
- Horn AG, Falls JB (1996) Categorisation and the design of signals: the case of song repertoires. In: Miller EH (ed) Ecology and evolution of acoustic communication in birds. Cornell University Press, Ithaca, N.Y., pp 121–136
- Horning CL, Beecher MD, Stoddard PK, Cambell SE (1996) Song perception in the song sparrow: importance of different parts of the song in song type classification. *Ethology* 94:46–58
- Jones AE, Cate Ct, Bijleveld CCJH (2001) The interobserver reliability of scoring sonagrams by eye: a study on methods, illustrated on zebra finch songs. *Anim Behav* 62:791–801
- Krebs JR, Kroodsma DE (1980) Repertoires and geographical variation in birdsong. *Adv Study Behav* 11:143–177
- Kroodsma DE (1982) Song repertoires: problems in their definition and use. In: Kroodsma DE, Miller EH (eds) Acoustic communication in birds. Academic Press, New York, pp 125–146
- Kroodsma DE, Parker LD (1977) Vocal virtuosity in brown thrasher. *Auk* 94:783–785
- Lampe HM, Espmark YO (1987) Singing activity and song pattern of the redwing *Turdus iliacus* during the breeding season. *Ornis Scand* 18:179–185
- Lampe HM, Espmark YO (1994) Song structure reflects male quality in pied flycatchers, *Ficedula hypoleuca*. *Anim Behav* 47:869–876
- Lebreton J-D, Burnham KP, Clobert J, Anderson DR (1992) Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecol Monogr* 62:67–118
- Lee SM, Chao A (1994) Estimating population size via sample coverage for closed capture-recapture models. *Biometrics* 50:88–97
- Leitner S, Voigt C, Gahr M (2001) Seasonal changes in the song pattern of the non-domesticated island canary (*Serinus canaria*), a field study. *Behaviour* 138:885–904
- Lessells CM, Boag PT (1987) Unrepeatable repeatabilities: a common mistake. *Auk* 104:116–121
- Nemeth E (1996) Different singing styles in mated and unmated reed buntings *Emberiza schoeniclus*. *Ibis* 138:172–176
- Nichols JD, Conroy MJ (1996) Estimation of species richness. In: Wilson DE, Cole FR, Nichols JD, Rudran R, Foster MS (eds) Measuring and monitoring biological diversity: standard methods for mammals. Smithsonian Institution Press, Washington D.C., pp 227–234
- Nichols JD, Boulinier T, Hines JE, Pollock KH, Sauer JR (1998) Inference methods for spatial variation in species richness and community composition when not all species are detected. *Conserv Biol* 12:1390–1398
- Norris JL, Pollock KH (1995) A capture-recapture model with heterogeneity and behavioural response. *Environ Ecol Stat* 2:305–313
- Norris JL, Pollock KH (1996) Non-parametric MLE under two closed capture-recapture models with heterogeneity. *Biometrics* 52:639–649
- Otis DL, Burnham KP, White GC, Anderson DR (1978) Statistical inference from capture data on closed animal populations. *Wildl Monogr* 62:1–135
- Pledger S (2000) Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics* 56:434–442
- Pollock KH, Otto MC (1983) Robust estimation of population size in closed animal populations from capture recapture experiments. *Biometrics* 39:1035–1049
- Pollock KH, Nichols JD, Brownie C, Hines JE (1990) Statistical inference for capture-recapture experiments. *Wildl Monogr* 107:1–97
- Rexstad E, Burnham KP (1991) User's guide for interactive program CAPTURE. Abundance estimation of closed animal population. Colorado State University, Fort Collins, Co, USA
- Riebel K, Slater PJB (1999) Song type switching in the chaffinch, *Fringilla coelebs*: timing or counting? *Anim Behav* 57:655–661
- Riebel K, Slater PJB (2000) Testing the flexibility of song type bout duration in the chaffinch, *Fringilla coelebs*. *Anim Behav* 59:1135–1142
- Rocha EPC, Danchin A, Viari A (1999a) Analysis of long repeats in bacterial genomes reveals alternative evolutionary mechanisms in *Bacillus subtilis* and other competent prokaryotes. *Mol Biol Evol* 16:1219–1230
- Rocha EPC, Danchin A, Viari A (1999b) Functional and evolutionary roles of long repeats in prokaryotes. *Res Microbiol* 150:725–733
- Saino N, Galeotti P, Sacchi R, Møller AP (1997) Song and immunological conditions in male barn swallows (*Hirundo rustica*). *Behav Ecol* 8:364–371
- Searcy W, Nowicki S, Peters S (1999) Song types as fundamental units in vocal repertoires. *Anim Behav* 58:37–44
- Slater PJB (1981) Chaffinch song repertoires—observations, experiments and discussion of their significance. *Z Tierpsychol* 56:1–24
- Titus RC (1998) Short-range and long-range songs: Use of two acoustically distinct song classes by dark-eyed juncos. *Auk* 115:386–393
- Titus RC, Ketterson ED, Nolan V (1997) High testosterone prior to song crystallization inhibits singing behavior in captive yearling dark-eyed juncos (*Junco hyemalis*). *Horm Behav* 32:133–140
- Van Duyse E, Pinxten R, Eens M (2002) Effects of testosterone on song, aggression, and nestling feeding behavior in male great tits, *Parus major*. *Horm Behav* 41:178–186
- Wildenthal JL (1965) Structure in primary song of the mockingbird (*Mimus polyglottos*). *Auk* 82:161–189
- Williams BK, Nichols JD, Conroy MJ (2002) Analysis and management of animal populations. Academic Press, New York