

RESEARCH ARTICLE

The assessment of biases in the acoustic discrimination of individuals

Pavel Linhart^{1,2*}, Martin Šálek^{3,4}

1 Ethology Department, Institute of Animal Science, Praha Uhřetěves, Czech Republic, **2** Department of Behavioural Ecology, Faculty of Biology, Adam Mickiewicz University, Poznań, Poland, **3** Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic, Brno, Czech Republic, **4** Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Praha, Czech Republic

* pavel.linhart83@gmail.com



Abstract

Animal vocalizations contain information about individual identity that could potentially be used for the monitoring of individuals. However, the performance of individual discrimination is subjected to many biases depending on factors such as the amount of identity information, or methods used. These factors need to be taken into account when comparing results of different studies or selecting the most cost-effective solution for a particular species. In this study, we evaluate several biases associated with the discrimination of individuals. On a large sample of little owl male individuals, we assess how discrimination performance changes with methods of call description, an increasing number of individuals, and number of calls per male. Also, we test whether the discrimination performance within the whole population can be reliably estimated from a subsample of individuals in a pre-screening study. Assessment of discrimination performance at the level of the individual and at the level of call led to different conclusions. Hence, studies interested in individual discrimination should optimize methods at the level of individuals. The description of calls by their frequency modulation leads to the best discrimination performance. In agreement with our expectations, discrimination performance decreased with population size. Increasing the number of calls per individual linearly increased the discrimination of individuals (but not the discrimination of calls), likely because it allows distinction between individuals with very similar calls. The available pre-screening index does not allow precise estimation of the population size that could be reliably monitored. Overall, projects applying acoustic monitoring at the individual level in population need to consider limitations regarding the population size that can be reliably monitored and fine-tune their methods according to their needs and limitations.

OPEN ACCESS

Citation: Linhart P, Šálek M (2017) The assessment of biases in the acoustic discrimination of individuals. *PLoS ONE* 12(5): e0177206. <https://doi.org/10.1371/journal.pone.0177206>

Editor: Tony Robillard, Museum National d'Histoire Naturelle, FRANCE

Received: September 5, 2016

Accepted: April 24, 2017

Published: May 9, 2017

Copyright: © 2017 Linhart, Šálek. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: PL was supported by the Czech Science Foundation (GA14-27925S), Ministry of Agriculture of Czech Republic (MZERO0716), and National Science Centre, Poland (2015/19/P/NZ8/02507); MŠ was funded by the Academy of Sciences of the Czech Republic (RVO68081766). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Introduction

Monitoring animals is a crucial activity for ecological, behavioural, and conservation science. There is now a growing interest in acoustic monitoring as an alternative or complementary means of monitoring animals [1]. At present, affordable hardware and software products are available making the practical use of acoustic monitoring more accessible [2]. The range of

Competing interests: The authors have declared that no competing interests exist.

considered acoustic monitoring applications ranges from detection of species presence, number and density of individuals of particular species and their activity in time and space to the assessment of diversity and health of whole ecosystems [1,3,4].

Many studies across the various taxa have demonstrated that, vertebrates universally have individually distinct vocalizations [5–12]. In other words, we can find one or more features in their vocalizations that are less variable within an individual than between individuals. In general, the individual distinctiveness may result from the unique vocal tract anatomy [13] and / or from the presence of unique arbitrary elements or variants in the repertoire of an individual that is used as an “individual signature” [14]. Vocal traits often vary along with physical or behavioural conditions of the individual. On the other hand, true identity signals should remain unaltered along significant time scales [15]. There are studies documenting the long-term stability of individual vocal traits in several bird and mammal species [16–18]. Thus, it is possible not only to discriminate between individuals but also to identify them in subsequent time periods [19]. Therefore, individual variation in vocalizations could in principle be used for the detailed and long-term acoustic monitoring of particular individuals. Several studies have shown that the acoustic identification of individuals could be a feasible and valuable tool [20–22], but it is still unclear what are potential biases associated with the particular study methods, design and sampling.

Any recognition task will be limited by the means and principles it uses [23]. The recognition of individuals usually involves two basic steps: 1) extraction of individually distinct features in calls and building a discrimination model and 2) attribution of new call samples to individuals using the discrimination model and evaluation of the discrimination model. While the performance of different classification methods has been compared and discussed before [24,25], the drawbacks and benefits of different methods of feature extraction are less well known. Very often, studies have used measurements of very specific vocalization subunits as individual features [21,26] fine-tuned to a particular species. These measurements may work fine for a single species, but must be developed and tested again and again for each new species. Other studies have used the cross-correlation method, in which the whole spectrogram of a call is compared to spectrograms of the other calls from known individuals and the call is then attributed to the individual with the highest concordance between spectrograms [27]. Cross-correlation does not extract any individual acoustic features per se but practically uses each pixel in the spectrogram as the feature. Cross-correlation scores are based on complete call representations that involve both frequency and amplitude modulation patterns of calls and thus cross-correlation could be probably considered as the most detailed method of call description. Further other studies have focused on more general properties of vocalizations, such as the distribution of the frequency spectrum, the distribution of formants and extracting Mel-frequency cepstral coefficients disregarding the specific composition of call / song subunits [28–30]. Such general approaches might have greater application potential across different species [31]. Few studies evaluated how the detail of the call description might influence the discrimination. An obvious assumption would be that the more detailed the call description the better the discrimination.

In real situations, if discrimination of individuals is used for the monitoring of individuals within a population, the number of monitored individuals would typically be relatively large. Studies investigating individual variation in vocalizations usually involved relatively small numbers of individuals, many of them including less than 20 individuals. Few studies with much larger samples of individuals have show that discrimination success decreases with population size [26,32] which is in accordance with theoretical assumptions [33]. Hence, it is important to understand how the population size being monitored may limit the accuracy of acoustic identification.

The discrimination of individuals is also limited by the quantity of sampling. Studies investigating individual variation typically use 10–20 calls per individual. Such numbers have been experimentally shown to be sufficient to assess amount of identity information in different species [34]. However, the number of required calls per individual will among other factors depend on external and internal factors affecting call consistency of a particular species. Therefore, more studies on additional species need to be carried out to understand how to scale sampling effort to achieve reasonable discrimination.

When preparing projects on acoustic monitoring of individuals, it might be very helpful to start with a small-scale, pre-screening pilot study to evaluate how much identity information is present and therefore, how many individuals could be discriminated in the species of interest and with the selected call features [34]. Researchers could then go on with a large-scale study if the results of the pre-screening were satisfying. Different measures / indices were used to assess the amount of identity information in vocalizations such as, for example, the score from discriminant analysis [26] or PIC—the potential of individual coding [35]. But only the Beecher's information criterion H_S [33,34] allows conversion to the number of potentially discriminable individuals. However, it is still not well known whether individuality measures, such as H_S in particular, are efficient for such pre-screening in different animals.

Owls (Strigiformes), including little owl, are excellent model organisms for acoustic monitoring because they rely on acoustic signals for the long-distance communication and hence are very vocal in different contexts. Moreover, several studies have demonstrated the short-term and long-term stability of individual call characteristics for a variety of owl species which is an important prerequisite for the efficient acoustic monitoring of individuals [6,36,37].

In this study, we assess several factors that could bias results of studies investigating potential for individual discrimination using an extensive sample of targeted recordings of the little owl *Athene noctua* individual males. We simulate effect of different methods and conditions on the discrimination performance at the level of calls and individuals to answer the following questions:

1. What is the difference in the discrimination performance among cross-correlation and two other frequently used methods of call description: call description by the fundamental frequency modulation, and by spectral features of vocalization?
2. How does the number of individuals being monitored affect discrimination performance and individuality index H_S ?
3. Could H_S be used to estimate the number of individuals which can be discriminated?
4. How does the number of calls available per male (i.e. sampling effort) affect the number of individual males that can be discriminated in a population?

Methods

Ethics statement

The study was done on places with unrestricted public access and on wild animals. Study was purely observational and non-invasive, therefore no special permits were required.

Study areas and species

The little owl is a non-migratory and sedentary nocturnal predator with stable long-term territories and low dispersal distances (< 15 km) of offspring [38]. We recorded territorial calls of

males that function both in territorial defense and mate choice [39,40] and other males can use them to distinguish their neighbours from strangers [41]. The species is strongly associated with open farmlands and its Western and Central European populations have steeply declined over the past 50 years, resulting in highly fragmented distribution and several local population extinctions [42–44].

The study was carried out in two Central European farmlands: 1) northern Bohemia, Czech Republic (50°23'N, 13°40'E) (CZ), 2) eastern Hungary, Hungary (47°33'N, 20°54'E) (HU). The mean population density of the little owl at the CZ site was 0.09 calling males per 10 km² and the population has experienced rapid population decline in recent years [44]. The mean population density of the little owl recorded at the HU site was 5.01 calling males per 10 km² [45] which is one of the highest population densities for this species in Central Europe [38]. The little owls in both study areas bred within the human settlements such as residential buildings and farmsteads [45,46].

Acoustic recording and analyses

Territorial calls [38] of each male were recorded for three minutes after a short playback provocation (≤ 1 min) inside their territories from up to 50 m distance from the individuals. We used a PMD660 solid-state recorder (sampling frequency 44 100 Hz, no compression) and a Sennheiser ME67 directional microphone to record the calls. Each recording contained calls of one focal male. The recordings were made during comparable, favourable meteorological conditions (without strong wind or precipitation), from sunset until midnight between March and April of 2013–2014. This period covered the mating season. The period and the time of the day for recording were selected with regard to the peak in vocal activity of little owls both within a day and within a season [47]. The recordings were band-pass filtered (500Hz–2000Hz) and down-sampled to 4000Hz sampling frequency prior to analyses as the fundamental frequency of calls was never below 500Hz nor exceeded 2000Hz (the minimum frequency of calls: mean \pm SD = 776 \pm 98 Hz; the maximum frequency of calls: mean \pm SD = 1668 \pm 272 Hz). Analyses were done in Avisoft SASLab Pro (Reimund Specht, Berlin). In all cases, spectrograms were generated with following settings: FFT-length was set to 512 points, the Flat Top window function was used, frame size was set to 100%, and window overlap was set to 93.75%.

Call description methods

We analysed calls from a subset of 54 males for which we had more than 20 calls each (20–41 calls per individual, mean \pm SD = 26.9 \pm 6.0) with good recording quality (14 individuals came from the CZ population, 40 individuals from the HU population). There were no differences in the spectral features or the frequency modulation of calls between the two populations (spectral features: MANOVA: Wilks = 0.80, P = 0.138, frequency modulation: MANOVA: Wilks = 0.88, P = 0.882). Hence, we pooled calls from the two populations for all analyses. Territorial calls were described based on the three approaches presented in Fig 1.

The first approach was based on the spectral features of the entire call (Fig 1b). We measured dominant frequency (dF, frequency of highest amplitude on the spectrum), frequencies at the three quartiles of amplitude distribution (q25, q50 and q75, below which lie respectively 25, 50 and 75% of the energy of the call) and minimum and maximum frequencies at -25dB relative to the call peak amplitude (minF, maxF, these two values give approximate range of fundamental frequency). Threshold of -25dB relative to the call peak amplitude was selected for two reasons: 1) setting the threshold makes measurements comparable between calls with variable absolute amplitudes and 2) the specific threshold value was selected based on „try and

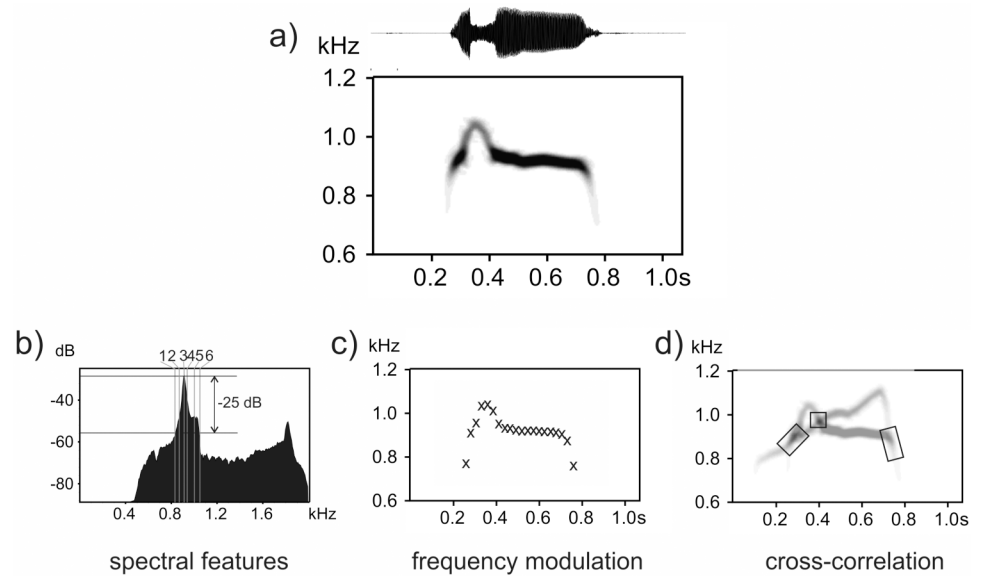


Fig 1. Illustration of little owl call and three methods used for the call description. Example the single territorial call of the little owl male (spectrogram and oscillogram, a), and an illustration of the three call description methods: b) description of call spectral features (1 = minF, 2 = q25, 3 = dF, 4 = q50, 5 = q75, 6 = maxF); c) description of call frequency modulation; and d) cross-correlation of calls (rectangles indicate cross-correlating segments between two displayed calls). Spectrogram settings: FFT-length = 512, window type = Flat Top, window overlap = 93.75%.

<https://doi.org/10.1371/journal.pone.0177206.g001>

error” to ensure that it was as close as possible to the minimum and the maximum fundamental frequency of the call but was within the call frequency range in all samples. This approach might be suitable in cases where the modulation of fundamental frequency differs between utterances, for example, in species which do not have a constant number of call elements, in which element types differ in a call sequence, or have noisy calls without clear fundamental frequency. It also can be used in species with complex songs [48]. The same set of features, in general, can be used across different species.

The second approach was based on the description of fundamental frequency modulation (Fig 1c). In this case, we took measurements of fundamental frequency at 20 measuring points (F1 –F20) evenly spaced throughout the duration of calls. Because discrimination based on 20 measuring points was not substantially better (Linhart, unpublished) we mostly used only 10 measuring points so that we took every second measuring point from the original 20 measuring points (see S1 Fig for the representations of F0 modulation and its variation in all 54 males). We used description based on 10 measuring points in all analyses with the exception of the analysis of H_S as a predictor of the discrimination performance (see below). The spectral features as well as the modulation of F0 were measured using the ‘Automatic parameter measurement’ tool in Avisoft SAS Lab (Reimund Specht, Berlin). Call duration was also measured in both cases.

In the third case, each call spectrogram was cross-correlated to the spectrograms of all other calls. We used the “Scan for template spectrogram patterns” function in Avisoft SASLab Pro. Settings were: high-pass cutoff frequency = 500Hz; low-pass cutoff frequency = 2000Hz; maximum frequency deviation = 50 Hz. This function returns cross-correlation scores between the template spectrogram and selected files. Each call was successively used as a template and was cross-correlated to all other calls so that we obtained a matrix of cross-correlation scores including all pair-wise combinations of calls in our dataset (Fig 1d).

Statistical analyses

General approach. In cases of spectral features and frequency modulation, we used linear discriminant analysis (LDA) with the leave-one-out cross validation to assign calls to individuals. Discriminant analyses were performed in R using the 'lda' function in the MASS package. We used leave-one-out cross-validation because the results were comparable to those obtained with generally stricter 2-fold cross-validation in the pilot test (see S2 Fig). Prior probability was set equal to each individual (computed as: 1 / number of individuals in a model).

To assign calls to individuals in the case of cross-correlation, we used the matrix of cross-correlation scores for each pair-wise combination of calls. The scores of calls belonging to the same male were then averaged and the call was assigned to the individual with highest average cross-correlation score.

Discrimination at the call and individual level. We report discrimination performance at two levels: level of call and level of individual. Similar studies report discrimination performance at the call level only [26,32]; this is equivalent to the frequently reported percentage of calls assigned to the correct individual by LDA. Performance at the level of individual can be easily derived from discrimination performance at the level of calls assuming that the whole set of calls (a calling bout) belongs to a single individual (e.g. when doing targeted recording of a single bird in sight). Although, some calls from the calling bout might be misattributed to other individuals, majority of the calls should be attributed to the correct individual. Therefore, we attributed the whole set of calls to an individual to whom the most of the calls from the set were assigned to (majority criterion, see Fig 2 for an example). Further, we take 90% of correctly discriminated individuals as a standard for acceptable discrimination at the individual level as this is comparable to the results from visual discrimination based on colour rings [49].

Call description method and population size to be monitored. We used custom built R scripts to simulate the effect of increasing population size (increasing number of individuals in LDA) and how it affects discrimination. We started by including calls from two randomly selected males in the LDA. Discrimination performance was evaluated (proportion of correctly identified calls and males). Then in each subsequent step another randomly selected male was

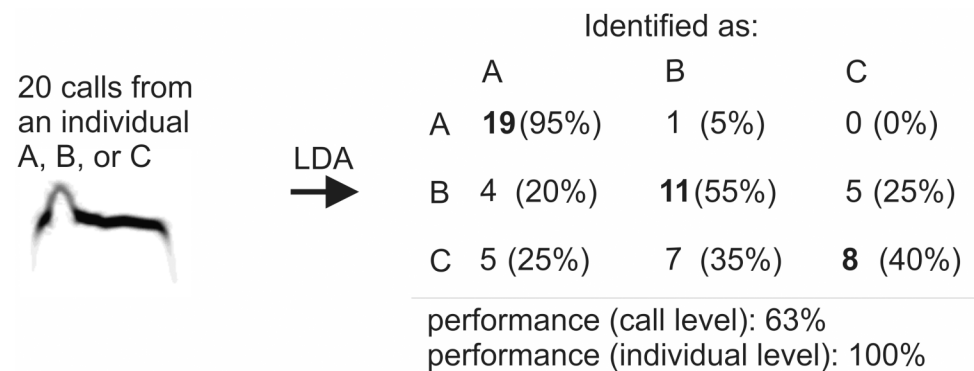


Fig 2. Relationship between the discrimination performance at the call and at the individual level. In this hypothetical example, calling bouts of 20 calls each from individuals A, B, and C are attributed to three individuals by linear discriminant analysis (LDA). Rows represent to which individual calls belonged to and columns represent to which individual calls were assigned to by LDA. Diagonal represents calls that were attributed to correct individuals. There is 100% discrimination success at the individual level because all three call sets were assigned to correct individual based on majority criterion. Even for C, the set of 20 calls would be correctly identified as belonging to individual C as majority of the calls (40%) were assigned to C. On the other hand, discrimination performance at the call level would be only 63% (overall percentage of correctly assigned calls).

<https://doi.org/10.1371/journal.pone.0177206.g002>

added to LDA model and the performance was evaluated until all 54 males were included in the LDA model. In the case of cross-correlation, the procedure was similar but calls were assigned based on cross-correlation scores. The whole run was repeated 20 times to simulate different combinations of individuals. We did not test for a statistical significance of differences in the performance of the methods explicitly. The performance of the three methods was compared graphically using average performance and confidence intervals.

Pre-screening of discrimination performance. Beecher’s information statistic H_S [33] is a stereotypy index commonly used to estimate the potential of the particular trait to signal individual identity. Higher values of H_S indicate greater potential to encode individual identity and are associated with better discrimination in LDA [33,50]. As in case of LDA, H_S was computed for sequentially increasing number of randomly chosen individuals from 2–54 males repeated 20 times, but only for the frequency modulation (10 measuring points). H_S was computed using the approach and the formulas from the previous studies [33,34]. First, we subjected the original acoustic variables (here F1–F10) to Principal Component Analysis (PCA) [33]. Original acoustic variables were scaled to zero mean and unit variance for PCA. For each of the resulting principle components (PC), we calculated its individual identity information content H_i :

Eq 1:

$$H_i = \log_2 \sqrt{\frac{F + n - 1}{n}} \tag{1}$$

where F is the F-statistic from an ANOVA with the particular PC entered as the dependent variable and the individual as the independent variable, n is the number of individual animals in the sample. Significant as well as non-significant F-values were used. The amount of individual identity information in the whole signal H_S is subsequently computed simply by summing identity information across all principle components:

Eq 2:

$$H_s = \sum H_i \tag{2}$$

The estimation of the number of individuals possible to discriminate was computed using another equation used by Pollard et al. [34] that follows from earlier equations used by Beecher [51]:

Eq 3:

$$N = P \times 2^{H_s} \tag{3}$$

where N is the number of individual animals distinguishable and P is the probability that a target individual’s signature is not held by another individual in the group. For monitoring purposes, we aim for individual traits that will provide a perfect, non-ambiguous identification ($P = 1$). However, this is rarely the case. Precision of the identification is not perfect even when using colour rings, so we set P to 0.9 in this study, which is comparable to the colour ring identification [49], a classical method to discriminate individual birds.

We were further interested in whether we could use H_S to pre-screen the best combination of acoustic parameters for call discrimination and to estimate the number of males that it would be possible to monitor. Therefore, H_S was computed for 23 LDA models that differed in how many and which measuring points (F1 –F20) were included (S1 Table). Each different model represents different amount of identity information available. This has been done for a full set of 54 males $H_S(54)$ as well as for a subpopulation of 10 individuals $H_S(10)$ (average from 20 random selections). We used the Spearman’s rank correlation and the linear

regression to test associations between: $H_S(10)$ and $H_S(54)$, $H_S(10)$ and the discrimination performance, the number of discriminable males estimated based on $H_S(10)$ and the number of males discriminated. This has been done to confirm, respectively, that H_S computed from the limited subset of males is closely related to H_S computed from the full set of males, that greater H_S is associated with higher discrimination performance and to see how accurately we can estimate the number of males being discriminated based on H_S . Linear regression was used when testing for association between the estimated and real number of discriminated individuals. Here, the linear relationship was expected because in an ideal situation, the real number of discriminated individuals should be equal to the estimate.

Number of calls and discrimination performance. The effect of an increasing number of calls available for LDA on classification success was also assessed using simulations. This was done again only for the frequency modulation. First we used all 54 males and increased the number of available calls from 2 to 20 to see how this affected discrimination performance. Again, we used 20 repetitions to simulate different call combinations. Finally, we combined both scenarios and simulated the effect of population size and number of calls simultaneously. We used 2 calls per male and increased the number of males from 2–54 (20 repetitions). We noted the average number of males for which the discrimination of individuals dropped under 90%, or where the overall discrimination of calls was lower than 65% (the worst documented call discrimination leading to more than 90% males correctly discriminated in our results), and we took it as a population size that could be reliably monitored for the particular number of calls per male. In subsequent steps this procedure was repeated with an increased number of calls per individual until 20 calls per individual were in the model.

Results

Call description method and population size

Discrimination performance at the level of calls. Overall, the discrimination performance was high and clearly exceeded discrimination expected by chance (discrimination expected by chance ranged from $1 / 2 = 50\%$ for 2 males; to $1 / 54 = 1.9\%$ for 54 males). The discrimination performance decreased steadily with an increasing number of individuals and ranged from 95% to 57% (Fig 3a). When all 54 males were included, discrimination based on

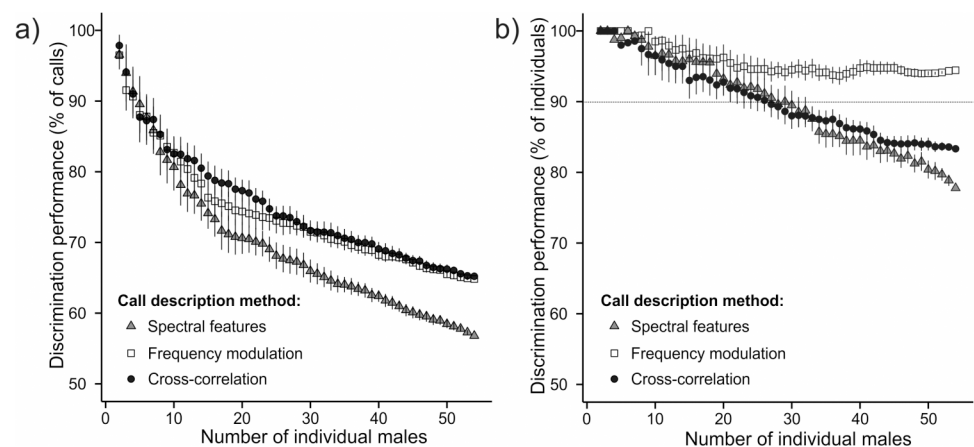


Fig 3. Effect of increasing number of individuals on discrimination performance. Effect of increasing number of individuals on discrimination performance at the level of calls (a) and at the level of individuals (b) for the three call description methods.

<https://doi.org/10.1371/journal.pone.0177206.g003>

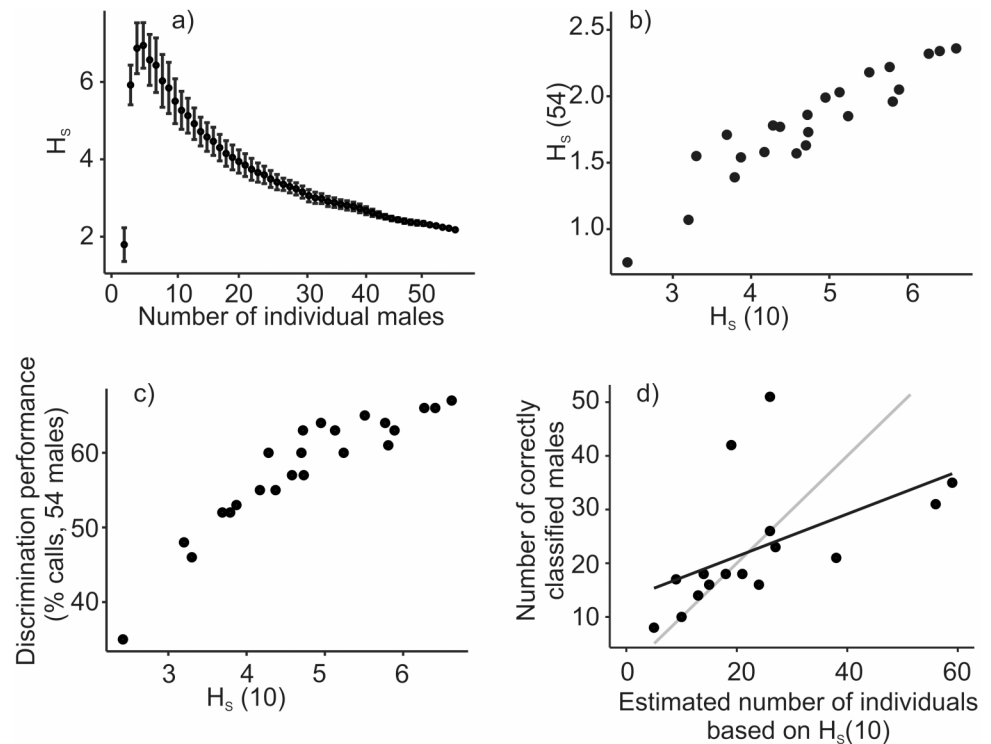


Fig 4. Relationship between the H_S and population size to be monitored. (a) H_S as a function of the number of individuals in the sample. (b) Relationship between average H_S computed from subsample of 10 random individuals $H_S(10)$ or full sample of 54 males $H_S(54)$. (c) Relationship between H_S and call discrimination performance. (d) Relationship between the estimated and real number of discriminated individuals. Grey line illustrates $y = x$ line for ideal estimates. H_S in (b), (c), and (d) was computed for 23 discrimination models that differed in how many and which measuring points (F1 –F20) were included (S1 Table).

<https://doi.org/10.1371/journal.pone.0177206.g004>

the cross-correlation scores performed best with a 65.2% success rate. However, the performance was similar in the case of LDA based on the frequency modulation (64.8%). LDA based on the spectral features performed the worst (56.8%).

Discrimination performance at the level of individuals. When we considered the performance of the three methods in the classification of individuals, we surprisingly found differing results. In this case, discrimination based on the cross-correlation scores (83.3%) performed better than LDA based on the spectral features (77.8%), and the LDA based on the frequency modulation was substantially better than other two methods with a 94.4% classification success (Fig 3b). Interestingly, there was a decrease in the performance for the LDA with the spectral features as well as in cross-correlation, but not in LDA with the frequency modulation which stabilized at about 95% of correctly identified males. On average, it would be possible to monitor 26 males with the cross-correlation, 27 males with the spectral features, but more than 54 males with the frequency modulation with 90% accuracy.

Pre-screening of discrimination performance

H_S in our study varied considerably with an increasing number of individuals (Fig 4a). An increasing number of individuals lead first to a steep increase in H_S ; it reached a peak at 5 individuals ($H_S = 6.94$) and then gradually decreased with each additional individual ($H_S = 2.18$ for 54 individuals). These H_S values would indicate a very wide and imprecise range of

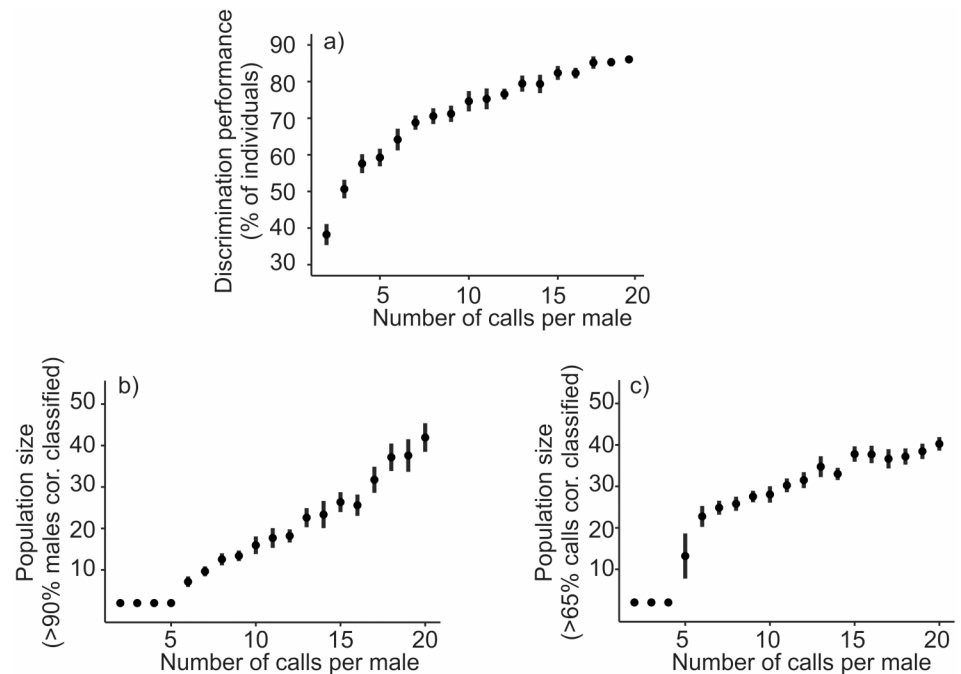


Fig 5. Effect of number of calls per male available on the discrimination performance. (a) Changes in performance with increasing number of calls available for discriminant function (for all 54 males). (b) Population size to be monitored if 90% individuals are to be classified correctly. (c) Population size to be monitored if 65% of calls are to be identified correctly.

<https://doi.org/10.1371/journal.pone.0177206.g005>

estimates of discriminable individuals—from 4 to 111 individuals could be discriminated assuming a 90% accuracy of recognition depending on how many individuals were sampled to calculate H_S .

$H_S(10)$ and $H_S(54)$ of the 23 models that differed in amount of identity information presented were positively correlated (Spearman rank correlation, $R = 0.93$, $P < 0.001$, Fig 4b). This shows, that variable sets having high identity information could be, on average, estimated by using only a subset of 10 individuals despite the fact that the absolute values of H_S change substantially with the number of individuals included. Further, Average $H_S(10)$ values were positively correlated with the performance of call discrimination for all 54 males (Spearman rank correlation, $R = 0.93$, $P < 0.001$, Fig 4c). Finally, the number of males estimated to be discriminated based on $H_S(10)$ was significantly positively associated with the number of males correctly discriminated in a full set of 54 males (linear regression: $F_{1,14} = 5.32$, adjusted $R^2 = 0.22$, $P = 0.037$, Fig 4d).

Number of calls and discrimination performance

The number of calls that were available for building a discriminant function affected call discrimination performance (Fig 5a). Performance increased steeply between 2–9 calls (72% correct at 9 calls) and then continued to increase up to 20 calls per male (90% correct) without reaching a stable plateau.

We further evaluated how an increasing number of calls influences how many individuals can be monitored with 90% precision. With less than 5 calls per individual, the 90% precision was never achieved (Fig 5b). Interestingly, in contrast to call discrimination, from 5–20 calls there was a steady, seemingly linear increase in the size of population that could be reliably

monitored with an increasing number of calls per male available. We, therefore, also tried to fix the overall call performance at at least 65% (lowest call discrimination performance documented in our analyses still leading to > 90% of individuals identified correctly) equivalently to fixing the performance at the individual level in the previous analysis (Fig 5c). In this case, there was a huge increase in the number of males from 4 (call discrimination never reached 65% or more) to 6 calls (call discrimination better than 65% for 23 males), a very slow increase with further added calls and no increase at all from c.a. 15 calls per male. This indicates that as few as 6–15 calls might be enough for correct call discrimination, but correct call discrimination is not sufficient for correct individual discrimination which in our case always benefits from adding more calls per individual.

Discussion

We found that discrimination performance decreased with an increasing number of individual males to be discriminated. Discrimination at the level of calls and at the level of individuals showed substantial discrepancies regarding the choice of the best feature description method and regarding insights into optimum recording effort per male. LDA based on frequency modulation performed best for discrimination of individuals and could be used to monitor more than 54 males if more than 90% males needed to be correctly identified. We found that, contrary to the expectations, the H_5 individuality index changed profoundly with the number of individuals. Nevertheless, H_5 correlated well with the call discrimination performance and could be used as relative index of individuality within the studied system. Higher number of calls per male had an important positive effect on discrimination performance. Interestingly, in our case, a high number of calls was not that crucial for discrimination at the level of single calls, but rather for assigning a whole call sequence to an individual. Call inconsistency negatively affected discrimination and was influenced by SNR. Internal factors also seem to cause part of call inconsistency.

Discrimination at the call and individual level

We show that slight differences at the level of call discrimination may have important consequences for discrimination at the level of individuals (misleading information about performance of the methods, choosing less efficient method, etc.). Researchers should take this into account when selecting the best method for individual recognition. Some studies have used quite strict rules and assigned a call sequence to an individual only if it received more than 50 or even 80 percent hits [22,52]. Our study shows that reliable recognition is possible even with a less strict rule, though at the expense of higher recording effort, i.e. recording more calls per individual.

Call description method and discrimination performance

We compared the performance of individual recognition based on three different methods. All three methods performed well above chance. Our results should be viewed as optimistic regarding the absolute values of discrimination performance because these might be lower if calls from different calling bouts had been used.

Cross-correlation has been suggested as best performing method for individual recognition [53,54]. In our study, cross-correlation performed slightly better than frequency modulation at the level of calls but fell behind at the level of the individual. Whether this is a general aspect of cross-correlation should be considered in future studies. Both methods, cross-correlation and frequency modulation, outperformed the LDA discrimination based on spectral features. This corresponds to the fact that owl hoots lack pronounced harmonics and formants. Hence, the

individual signature is likely to be conveyed by the frequency modulation. Description of the hoot frequency modulation is also commonly used in other studies investigating individual variation of owl hoots [6,55,56].

The three methods differed regarding the call description detail and specificity to the study system. The assumptions on how individuality is encoded in the call differ between the three methods. Cross-correlation might be considered as the most detailed method of call description because every single spectrogram point is considered to compute similarity. On the other hand, spectral features do provide only very general an uncomplete call description. It is, therefore, surprising that the performance of the two methods at the individual level was alike and relatively good: allowing discrimination of c.a. 30 males with 90% accuracy. Probably, good identity signals can be narrowed down to few parameters despite their complexity, so that they can enhance individual recognition and keep low processing demands at the same time [15,57]. To develop acoustic monitoring of individuals, researchers might benefit more from spending time to search for the best marker of identity among different vocalisation types rather than focusing on a single vocalization type.

Pre-screening of discrimination performance

Because individual discrimination can be compromised in large populations, it is necessary to make use of pre-screening procedures to see whether the species of interest and intended methods will give the appropriate results in cases of large-scale application [34]. Beecher's informative criterion H_S is the only individuality metric currently available that allows a direct conversion of an individuality index into a number of discriminable animals. We found that on average H_S computed for a subset of 10 males correlated well with the H_S in a complete set of 54 males and even with call classification success. This is in agreement with a previous study [33]. However, the relationship between the actual number of males correctly classified and that estimated from H_S was not very tight. Hence, we argue the H_S gives a good relative measure of individuality but cannot be used to estimate the size of the population that can be monitored.

Moreover, we found that H_S changes very markedly with population size although H_S , unlike the LDA classification success scores, has been suggested to be independent of sampling [33]. The effect of the number of individuals on H_S , though small, has also been found previously [34], suggesting that comparisons of H_S values from different studies might be problematic. In the original study, there was not apparent effect of number of individuals on H_S [33]. Studies might have underestimated this effect due to the numbers of individuals used in previous studies might be drawn from the two sides of the H_S peak (Fig 4a). For example, in case we would include 3–10 individuals, we would likely not detect any linear relationship between H_S and number of individuals, while including 10–20 individuals into analysis would probably result in negative relationship between the two. Alternatively, the relationship between H_S and number of individuals (Fig 4a) does not represent general pattern and could be specific to our study system. Why H_S first rises and then falls again and whether it is a general pattern needs to be explained in further studies. But it is possible that the rise reflects the rapid initial expansion of acoustic space each time the new individual is included (i.e. within one dimension the variance between individuals increases while the variance within individuals remains similar).

Number of calls and discrimination performance

We show that discrimination improves with the number of calls available per individual which is in accordance with a previous study [34]. The previous and this study (Fig 5c) both agree that relatively small number of calls is sufficient to assess the amount of individual information

in the calls. However, our study shows that the population size that can be reliably monitored increases approximately linearly with the number of calls available and that acoustic monitoring programs would likely benefit from increased recording effort. Many calls are not crucial at the level of building discrimination function, because the within-individual variation in calls is low. On the other hand, large number of calls becomes necessary for reliable attribution of those calls to a specific individual if the between-individual variation in calls is not high enough to allow for unambiguous discrimination.

Conclusions

To conclude, future studies comparing methods of individual discrimination should consider to implement metrics of performance at the level of individuals rather than at the call level only. If researchers plan to individual acoustic monitoring on large scales, they can select the best performing method of call description by pre-screening a limited number of individuals. However, it is not possible to safely estimate the population size for which that method would perform satisfactorily. For small populations, selection of the call description method might not be crucial and even very general methods could be useful. Large scale applications should benefit from collecting large number of calls per individual. Despite the fact that large number of calls per individual is not crucial for building discrimination model, high number of calls per individual is crucial to reliably attribute the sequence of calls to correct individual in larger populations.

An important finding of our study is that discrimination performance (percentage of correctly assigned calls or individuals) and H_S are influenced by sampling of the study. Therefore, they should not be directly compared between studies. Robust and accurate pre-screening techniques are currently lacking and should be developed in order to provide a tool to assess the degree of individuality in vocalizations and the efficiency of different methods for the acoustic individual discrimination and identification.

Supporting information

S1 Dataset. Spectral features and frequency modulation for analysed calls.
(XLSX)

S2 Dataset. Pair-wise cross-correlation scores for analysed calls.
(XLSX)

S1 Fig. Call spectrograms of 54 individual males.
(DOCX)

S2 Fig. Comparison of LDA performance with leave-one-out and split sample cross-validation.
(DOCX)

S1 Table. Overview of 23 different discrimination models based on F1–F20 measuring points.
(DOCX)

Acknowledgments

We would like to thank to the anonymous reviewers whose comments helped to improve the clarity and quality of the article. Marina Kipson, Monika Chrenková helped to collect data in

the field. Reimund Specht provided a great support regarding Avisoft SASLab Pro cross-correlation. Daniel Blumstein provided feedback on early draft of the manuscript.

Author Contributions

Conceptualization: PL MS.

Data curation: PL MS.

Formal analysis: PL.

Funding acquisition: PL MS.

Investigation: PL MS.

Methodology: PL MS.

Project administration: PL MS.

Resources: PL MS.

Supervision: PL MS.

Visualization: PL MS.

Writing – original draft: PL MS.

Writing – review & editing: PL MS.

References

1. Blumstein DT, Mennill DJ, Clemins P, Girod L, Yao K, Patricelli G, et al. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *J Appl Ecol.* 2011; 48: 758–767.
2. Wilson DR, Battiston M, Brzustowski J, Mennill DJ. Sound Finder: a new software approach for localizing animals recorded with a microphone array. *Bioacoustics.* 2014; 23: 99–112.
3. Merchant ND, Fristrup KM, Johnson MP, Tyack PL, Witt MJ, Blondel P, et al. Measuring acoustic habitats. *Methods Ecol Evol.* 2015; 6: 257–265. <https://doi.org/10.1111/2041-210X.12330> PMID: 25954500
4. Petrusková T, Pišvejcová I, Kinštová A, Brinke T, Petrusek A. Repertoire-based individual acoustic monitoring of a migratory passerine bird with complex song as an efficient tool for tracking territorial dynamics and annual return rates. *Methods Ecol Evol.* 2015; n/a–n/a.
5. Robertson BC. Vocal mate recognition in a monogamous, flock-forming bird, the silvereye, *Zosterops lateralis*. *Anim Behav.* 1996; 51: 303–311.
6. Delport W, Kemp AC, Ferguson JWH. Vocal identification of individual African Wood Owls *Strix woodfordii*: a technique to monitor long-term adult turnover and residency. *Ibis.* 2002; 144: 30–39.
7. Feng AS, Riede T, Arch VS, Yu Z, Xu Z-M, Yu X-J, et al. Diversity of the vocal signals of concave-eared torrent frogs (*Odontophrynus tormota*): evidence for individual signatures. *Ethology.* 2009; 115: 1015–1028.
8. Schneiderová I, Policht R. Alarm Calls of the European Ground Squirrel *Spermophilus Citellus* and the Taurus Ground Squirrel *S. Taurensis* Encode Information About Caller Identity. *Bioacoustics- Int J Anim Sound Its Rec.* 2010; 20: 29–43.
9. Amorim MCP, Simoes JM, Almada VC, Fonseca PJ. Stereotypy and variation of the mating call in the Lusitanian toadfish, *Halobatrachus didactylus*. *Behav Ecol Sociobiol.* 2011; 65: 707–716.
10. Antunes R, Schulz T, Gero S, Whitehead H, Gordon J, Rendell L. Individually distinctive acoustic features in sperm whale codas. *Anim Behav.* 2011; 81: 723–730.
11. Cinková I, Policht R. Contact Calls of the Northern and Southern White Rhinoceros Allow for Individual and Species Identification. *Plos One.* 2014; 9: e98475. <https://doi.org/10.1371/journal.pone.0098475> PMID: 24901244
12. Salmi R, Hammerschmidt K, Doran-Sheehy DM. Individual Distinctiveness in Call Types of Wild Western Female Gorillas. *Plos One.* 2014; 9: e101940. <https://doi.org/10.1371/journal.pone.0101940> PMID: 25029238

13. Taylor AM, Reby D. The contribution of source—filter theory to mammal vocal communication research. *J Zool.* 2010; 280: 221–236.
14. Janik VM, Sayigh LS. Communication in bottlenose dolphins: 50 years of signature whistle research. *J Comp Physiol -Neuroethol Sens Neural Behav Physiol.* 2013; 199: 479–489.
15. Tibbetts E, Dale J. Individual recognition: it is good to be different. *Trends Ecol Evol.* 2007; 22: 529–537. <https://doi.org/10.1016/j.tree.2007.09.001> PMID: 17904686
16. Feng J-J, Cui L-W, Ma C-Y, Fei H-L, Fan P-F. Individuality and Stability in Male Songs of Cao Vit Gibbons (*Nomascus nasutus*) with Potential to Monitor Population Dynamics. *PLoS ONE.* 2014; 9: e96317. <https://doi.org/10.1371/journal.pone.0096317> PMID: 24788306
17. Klenova AV, Volodin IA, Volodina EV. Examination of pair-duet stability to promote long-term monitoring of the endangered red-crowned crane (*Grus japonensis*). *J Ethol.* 2008; 27: 401–406.
18. Peake TM, McGregor PK, Smith KW, Tyler G, Gilbert G, Green RE. Individuality in Corncrake *Crex crex* vocalizations. *Ibis.* 1998; 140: 120–127.
19. Terry AM, Peake TM, McGregor PK. The role of vocal individuality in conservation. *Front Zool.* 2005; 2: 10. <https://doi.org/10.1186/1742-9994-2-10> PMID: 15960848
20. Adi K, Johnson MT, Osiejuk TS. Acoustic censusing using automatic vocalization classification and identity recognition. *J Acoust Soc Am.* 2010; 127: 874–883. <https://doi.org/10.1121/1.3273887> PMID: 20136210
21. Laiolo P, Vogeli M, Serrano D, Tella J. Testing acoustic versus physical marking: two complementary methods for individual-based monitoring of elusive species. *J Avian Biol.* 2007; 38: 672–681.
22. Terry AMR, McGregor PK. Census and Monitoring Based on Individually Identifiable Vocalizations: The Role of Neural Networks. *Anim Conserv.* 2002; 5: 103–111.
23. Janik VM. Pitfalls in the categorization of behaviour: a comparison of dolphin whistle classification methods. *Anim Behav.* 1999; 57: 133–143. <https://doi.org/10.1006/anbe.1998.0923> PMID: 10053080
24. Arriaga JG, Sanchez H, Hedley R, Vallejo EE, Taylor CE. Using Song to Identify Cassin's Vireo Individuals. A Comparative Study of Pattern Recognition Algorithms. In: Martínez-Trinidad JF, Carrasco-Ochoa JA, Olvera-Lopez JA, Salas-Rodríguez J, Suen CY, editors. *Pattern Recognition.* Springer International Publishing; 2014. pp. 291–300.
25. Kirschel ANG, Earl DA, Yao Y, Escobar IA, Vilches E, Vallejo EE, et al. Using Songs to Identify Individual Mexican Anthrush *Formicarius Moniliger*: Comparison of Four Classification Methods. *Bioacoustics.* 2009; 19: 1–20.
26. Xia C, Lin X, Liu W, Lloyd H, Zhang Y. Acoustic Identification of Individuals within Large Avian Populations: A Case Study of the Brownish-Flanked Bush Warbler, South-Central China. *PLoS ONE.* 2012; 7: e42528. <https://doi.org/10.1371/journal.pone.0042528> PMID: 22880018
27. Kennedy RAW, Evans CS, McDonald PG. Individual distinctiveness in the mobbing call of a cooperative bird, the noisy miner *Manorina melanocephala*. *J Avian Biol.* 2009; 40: 481–490.
28. Fox EJS. A new perspective on acoustic individual recognition in animals with limited call sharing or changing repertoires. *Anim Behav.* 2008; 75: 1187–1194.
29. Budka M, Osiejuk TS. Formant frequencies are acoustic cues to caller discrimination and are a weak indicator of the body size of corncrake males. *Ethology.* 2013; 119: 960–969.
30. Ptáček L, Machlica L, Linhart P, Jaška P, Müller L. Automatic recognition of bird individuals on an open set using as-is recordings. *Bioacoustics- Int J Anim Sound Its Rec.* 2016; 25: 55–73.
31. Cheng J, Sun Y, Ji L. A call-independent and automatic acoustic system for the individual recognition of animals: A novel model using four passerines. *Pattern Recognit.* 2010; 43: 3846–3852.
32. Budka M, Wojas L, Osiejuk TS. Is it possible to acoustically identify individuals within a population? *J Ornithol.* 2014; 156: 481–488.
33. Beecher M. Signaling systems for individual recognition—an information-theory approach. *Anim Behav.* 1989; 38: 248–261.
34. Pollard KA, Blumstein DT, Griffin SC. Pre-screening acoustic and other natural signatures for use in noninvasive individual identification. *J Appl Ecol.* 2010; 47: 1103–1109.
35. Lengagne T, Lauga J, Jouventin P. A method of independent time and frequency decomposition of bioacoustic signals: inter-individual recognition in four species of penguins. *Comptes Rendus Académie Sci Sér III Sci Vie.* 1997; 320: 885–891.
36. Lengagne T. Temporal Stability in the individual features in the calls of eagle owls (*Bubo bubo*). *Behaviour.* 2001; 138: 1407–1419.
37. Odom KJ, Slaght JC, Gutiérrez R J. Distinctiveness in the Territorial Calls of Great Horned Owls within and among Years. *J Raptor Res.* 2013; 47: 21–30.

38. Nieuwenhuysse DV, Génot J-C, Johnson DH. The Little Owl: Conservation, Ecology and Behavior of *Athene noctua*. Cambridge University Press; 2008.
39. Exo KM, Scherzinger W. Voice and inventory of call-notes of the little owl (*Athene noctua*): description, context and habitat adaptation. *Ecol Birds*. 1989; 149–187.
40. Jacobsen LB, Sunde P, Carsten R, Dabelsteen T, Thorup K. Territorial calls in the Little Owl (*Athene noctua*): spatial dispersion and social interplay of mates and neighbours. *Ornis Fenn*. 2013; 41–49.
41. Hardouin LA, Tabel P, Bretagnolle V. Neighbour—stranger discrimination in the little owl, *Athene noctua*. *Anim Behav*. 2006; 72: 105–112.
42. Zmihorski M, Altenburg-Bacia D, Romanowski J, Kowalski M, Osojca G. Long-term decline of the little owl (*Athene noctua Scop.*, 1769) in Central Poland. *Pol J Ecol*. 2006; 54: 321–324.
43. Šálek M, Schröpfer L. Population decline of the Little Owl (*Athene noctua*) in the Czech Republic. *Pol J Ecol*. 2008; 56: 527–534.
44. Šálek M. Dlouhodobý pokles početnosti sýčka obecného (*Athene noctua*) v jádrové oblasti jeho rozšíření v Čechách. *Sylvia*. 2014; 50: 2–11.
45. Šálek M, Chrenková M, Kipson M. High Population Density of Little Owl (*Athene noctua*) in Hortobagy National Park, Hungary, Central Europe. *Pol J Ecol*. 2013; 61: 165–169.
46. Šálek M, Chrenková M, Dobrý M, Kipson M, Grill S, Václav R. Scale-dependent habitat associations of a rapidly declining farmland predator, the Little Owl *Athene noctua*, in contrasting agricultural landscapes. *Agric Ecosyst Environ*. 2016; 224: 56–66.
47. Exo KM. Tagesperiodische Aktivitätsmuster des Steinkauzes (*Athene noctua*). *Vogelwarte*. 1989; 35: 99–114.
48. Weary D, Norris K, Falls J. Song Features Birds Use to Identify Individuals. *Auk*. 1990; 107: 623–625.
49. Milligan JL, Davis AK, Altizer SM. Errors Associated with Using Colored Leg Bands to Identify Wild Birds (Errores asociados con el uso de bandas coloreadas para las patas para identificar aves silvestres). *J Field Ornithol*. 2003; 74: 111–118.
50. Medvin MB, Stoddard PK, Beecher MD. Signals for parent-offspring recognition: a comparative analysis of the begging calls of cliff swallows and barn swallows. *Anim Behav*. 1993; 45: 841–850.
51. Beecher M. Signature Systems and Kin Recognition. *Am Zool*. 1982; 22: 477–490.
52. Fernandez-Juricic E, del Nevo AJ, Poston R. Identification of Individual and Population-Level Variation in Vocalizations of the Endangered Southwestern Willow Flycatcher (*empidonax Traillii Extimus*). *Auk*. 2009; 126: 89–99.
53. Lein MR. Song Variation in Buff-Breasted Flycatchers (*Empidonax fulvifrons*). *Wilson J Ornithol*. 2008; 120: 256–267.
54. Xia C, Huang R, Wei C, Nie P, Zhang Y. Individual identification on the basis of the songs of the Asian Stubtail (*Urosphena squameiceps*). *Chin BIRDS*. 2011; 2: 132–139.
55. Grava T, Mathevon N, Place E, Balluet P. Individual acoustic monitoring of the European Eagle Owl *Bubo bubo*. *Ibis*. 2008; 150: 279–287.
56. Rognan CB, Szewczak JM, Morrison ML. Vocal Individuality of Great Gray Owls in the Sierra Nevada. *J Wildl Manag*. 2009; 73: 755–760.
57. Wiley RH. Specificity and multiplicity in the recognition of individuals: implications for the evolution of social behaviour. *Biol Rev*. 2013; 88: 179–195. <https://doi.org/10.1111/j.1469-185X.2012.00246.x> PMID: 22978769