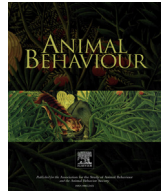




Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Data Set

A densely sampled and richly annotated acoustic data set from a wild bird population

Nilo Merino Recalde ^{a,*} , Andrea Estandía ^a , Loanne Pichot ^b, Antoine Vansse ^b , Ella F. Cole ^a , Ben C. Sheldon ^a ^a Edward Grey Institute, Department of Biology, University of Oxford, Oxford, U.K.^b École Normale Supérieure de Lyon, Lyon, France

ARTICLE INFO

Article history:

Received 30 June 2023

Initial acceptance 28 July 2023

Final acceptance 8 January 2024

Available online xxx

MS. number: 23-00346

Keywords:

animal culture
animal vocalization
bioacoustics
birdsong
open data

We present a high-resolution, densely sampled data set of wild bird songs collected over multiple years from a single population of great tits, *Parus major*, in the U.K. The data set includes over 1 000 000 individual acoustic units from 109 963 richly annotated songs, sung by more than 400 individual birds, and provides unprecedented detail on the vocal behaviour of wild birds. Here, we describe the data collection and processing procedures and provide a summary of the data. We also discuss potential research questions that can be addressed using this data set, including behavioural repeatability and stability, links between vocal performance and reproductive success, the timing of song production, syntactic organization of song production and song learning in the wild. We have made the data set and associated software tools publicly available with the aim that other researchers can benefit from this resource and use it to further our understanding of bird vocal behaviour in the wild.

© 2024 The Authors. Published by Elsevier Ltd on behalf of The Association for the Study of Animal Behaviour. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Despite a long history of scientific interest from disciplines as diverse as behavioural ecology, neurobiology and physiology, there is still much to learn regarding the evolution and function of animal vocalizations. Ongoing research covers a wide range of topics, including speech recognition and language evolution in humans, animal welfare and even fish vocal communication. The study of animal vocalizations offers valuable insights into the intricacies of social interactions and reproductive strategies. They frequently convey crucial information about an individual's condition and identity (Lehmann & Seufert, 2017; Linhart et al., 2019), the cohesion of social groups and the structure of social hierarchies (Bell et al., 2010; Engesser & Manser, 2022; Radford & Ridley, 2007). Additionally, animal vocalizations play a substantial role in the formation of social bonds, the selection of mates and the provision of parental care (Behr & von Helversen, 2004; Gerhardt, 1991; Pitcher et al., 2010; Roulin, 2001).

For those interested in social learning and cultural evolution, animal vocalizations, particularly those of birds, have long been a focus of research. This interest dates back at least to the pioneering work of Marler and Thorpe with chaffinches, *Fringilla coelebs*, and

white-crowned sparrows, *Zonotrichia leucophrys* (Marler, 1952; Marler & Tamura, 1962, 1964; Thorpe, 1958), which paved the way for what continues to be a thriving field today (see Mets & Brainard, 2019; Riebel et al., 2015; Williams & Lachlan, 2021; Youngblood & Lahti, 2022). In addition, and from a more mechanistic point of view, they offer a window into the physiological and neural mechanisms underlying vocal production and perception, as well as the consolidation of memories and motor coordination, to name but a few (Davenport & Jarvis, 2023).

Beyond their fundamental scientific importance, animal vocalizations have practical applications in various fields. For example, there is increasing recognition of their potential as a noninvasive tool for monitoring populations. By analysing entire soundscapes, researchers can gather crucial information about population dynamics, species distribution and the presence of rare or elusive species (Kahl et al., 2021; Sethi et al., 2020; Sugai et al., 2019). However, despite the growing interest in animal vocalizations and their potential applications, publicly available data from wild populations are still scarce, with the *xeno-canto* community science project as a prominent exception, focusing primarily on sparse recordings of most of the world's bird species rather than dense sampling of populations within the same species. This can severely limit researchers' ability to ask questions that require large data sets to answer, such as those about social learning, vocal

* Corresponding author.

E-mail address: nilo.recalde@biology.ox.ac.uk (N. Merino Recalde).<https://doi.org/10.1016/j.anbehav.2024.02.008>0003-3472/© 2024 The Authors. Published by Elsevier Ltd on behalf of The Association for the Study of Animal Behaviour. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).Please cite this article in press as: Merino Recalde, N., et al., A densely sampled and richly annotated acoustic data set from a wild bird population, *Animal Behaviour* (2024), <https://doi.org/10.1016/j.anbehav.2024.02.008>

development, large-scale cultural diversity and the syntactic structure of animal vocalizations (Aplin, 2019; Kollmorgen et al., 2020; Lachlan et al., 2018; Sainburg et al., 2019). Indeed, while controlled laboratory settings allow researchers to track vocal development and production in minute detail, it is much harder to obtain finely grained data from animals in their natural habitats. The process of collecting such data can be very demanding and requires significant time, technical expertise and resources: this includes both data collection itself and the subsequent processing of acoustic data files.

A second limitation arises after data have been collected, due to (1) researchers' understandable focus on specific, often narrowly defined questions, (2) practical constraints and (3) scientific cultural norms that have not encouraged data sharing. Combined, these factors often lead to a tendency of not publishing or only partially publishing the data collected during research. This lack of data sharing can hinder scientific progress and makes it difficult to reproduce research findings (Jenkins et al., 2023; Powers & Hampton, 2019; Reichman et al., 2011; Wilkinson et al., 2016); hence, we argue that there is great intrinsic value in publishing fully curated acoustic data sets. If this practice becomes widespread, it will allow scientists to explore a broader range of research questions, improve reproducibility and facilitate the validation of findings across different studies and populations (Hersh et al., 2023; Powers & Hampton, 2019).

In line with this perspective, we present a comprehensive data set of wild birdsongs recorded from a single population of great tits, *Parus major*, in Wytham Woods, Oxford, U.K. We collected 21 283 h of continuous recordings across 703 nesting sites over three spring seasons, which resulted in the annotation of over 1 100 000 notes or acoustic units from more than 100 000 songs (see below for definitions of these terms), sung by approximately 400 different male great tits. Among these birds, we have detailed information on the identity and life history of 242 individuals, including 50 that were recorded in multiple years. This information includes the time and location of breeding attempts, clutch size, number of fledglings, age of the bird and basic morphological traits. For birds born in the population (106, or 43% of the total), we also include details such as birthplace, postnatal dispersal distance, mother and social father.

To complement the song recordings, we have prepared extensive metadata for each of the more than 100 000 songs. This includes details such as the onset and offset times of each note within the song, a song type label and the time of recording. We also provide the time of the first song at dawn. Finally, we augment the data set by providing embeddings of each song, which are vector representations derived from a deep metric learning model specifically trained on this data set. These can be used to identify individuals and in tasks that require similarity judgements.

Great tit song has been the subject of extensive research activity (see, for example, Lambrechts & Dhondt, 1990; Lind et al., 1996; Ritschard et al., 2012; Rivera-Gutierrez, Matthysen et al., 2010; Rivera-Gutierrez, Pinxten et al., 2010; Rivera-Gutierrez et al., 2012; Rivera-Gutierrez et al., 2011; Slagsvold, Sætre, & Dale, 1994). Research conducted within the Wytham Woods population, in particular, has given rise to many influential ideas and insights into bird singing behaviour. These include investigations into neighbour interactions, song matching and the connection between song repertoires and reproductive success (McGregor & Krebs, 1989; McGregor et al., 1981; McGregor et al., 1983), the dynamics of song learning from neighbouring individuals and the acquisition of distinct song types (McGregor & Krebs, 1982, 1989), as well as the role of song repertoires in maintaining territories and reducing listener habituation (Krebs, 1976; Krebs et al., 1978), the functions

of dawn song (Kacelnik & Krebs, 1983; Mace, 1987), and the influence of spatial factors and movement on song culture (Fayet et al., 2014). We hope that this data set, which is, to the best of our knowledge, the largest publicly available collection of birdsongs from a single wild population, will contribute to that effort by providing valuable insights into a range of scientific questions, including behavioural repeatability and stability, links between vocal performance and reproductive success, the timing of song production, the syntactic organization of song production and song learning in the wild.

What follows is a detailed description of the data collection and curation process and the resulting data set, together with some discussion around potential uses of data presented in this format.

DATA COLLECTION

Study System and Fieldwork

Great tits are small, short-lived birds (average reproductive life span: 1.9 years) that sing acoustically simple yet highly diverse songs. During the breeding season, from March to June, Great tit pairs are socially monogamous and defend territories around their nests (Hinde, 1952). In Wytham Woods (51°46'N, 1°20'W), a population of these birds has been the focus of a long-term study since 1947 (Lack, 1964). Wytham Woods is a seminatural predominantly deciduous woodland that spans an area of approximately 385 ha and is surrounded by farmland. Most great tits in this population breed in nestboxes with known locations (see map in Fig. 1a), and the majority of individuals are marked with a unique British Trust for Ornithology (BTO) metal leg ring as nestlings or adults.

We collected data from late March to mid-May during the breeding seasons of 2020, 2021 and 2022. Every year, fieldworkers checked each of the 1018 nestboxes at least once a week before and during the egg-laying period, which typically lasts 1–14 days (Perrins, 1965), and recorded the identities of breeding males and females, the dates of clutch initiation and egg hatching, clutch size, and fledgling number and condition under standardized protocols. We found the first egg date by assuming that one egg is laid every day and counting back from the day of observation. In cases where we did not observe the chicks on the day of hatching, the actual hatching date was determined by assessing the weight of the heaviest chicks and extrapolating their age from established growth curves.

To record the vocalizations of male great tits, we took advantage of their behaviour during the reproductive period, when they engage in continuous singing near their nests at dawn before and during egg laying (Mace, 1987). Collectively, this vocal display is referred to as the dawn chorus and has been demonstrated to yield a reliable estimation of the song repertoire of individuals when recorded in full (Rivera-Gutierrez et al., 2012; Van Duyse et al., 2005). As soon as we suspected that a pair of great tits were using a nestbox based on nest-lining materials, egg size if present or other signs of activity, we deployed an autonomous sound recorder nearby. These recorders were placed on the trunk of the same tree or on a nearby tree, between 1 and 2 m above the ground and no more than 5 m away, depending on tree availability. We aimed to keep the recorder in a consistent position and orientation. The microphone pointed upwards and slightly away from the nestbox, in the same direction as the entrance hole. The birds sang close to the recorder and moved around. (We were not able to collect data on the bird's distance to the recorder, but the mean distance to the nestbox was 10 m in a different population studied by Halfwerk et al., 2012, which matches our anecdotal observations.) Although changes in amplitude due to distance and directionality impacted song selection, we did not observe any systematic bias.

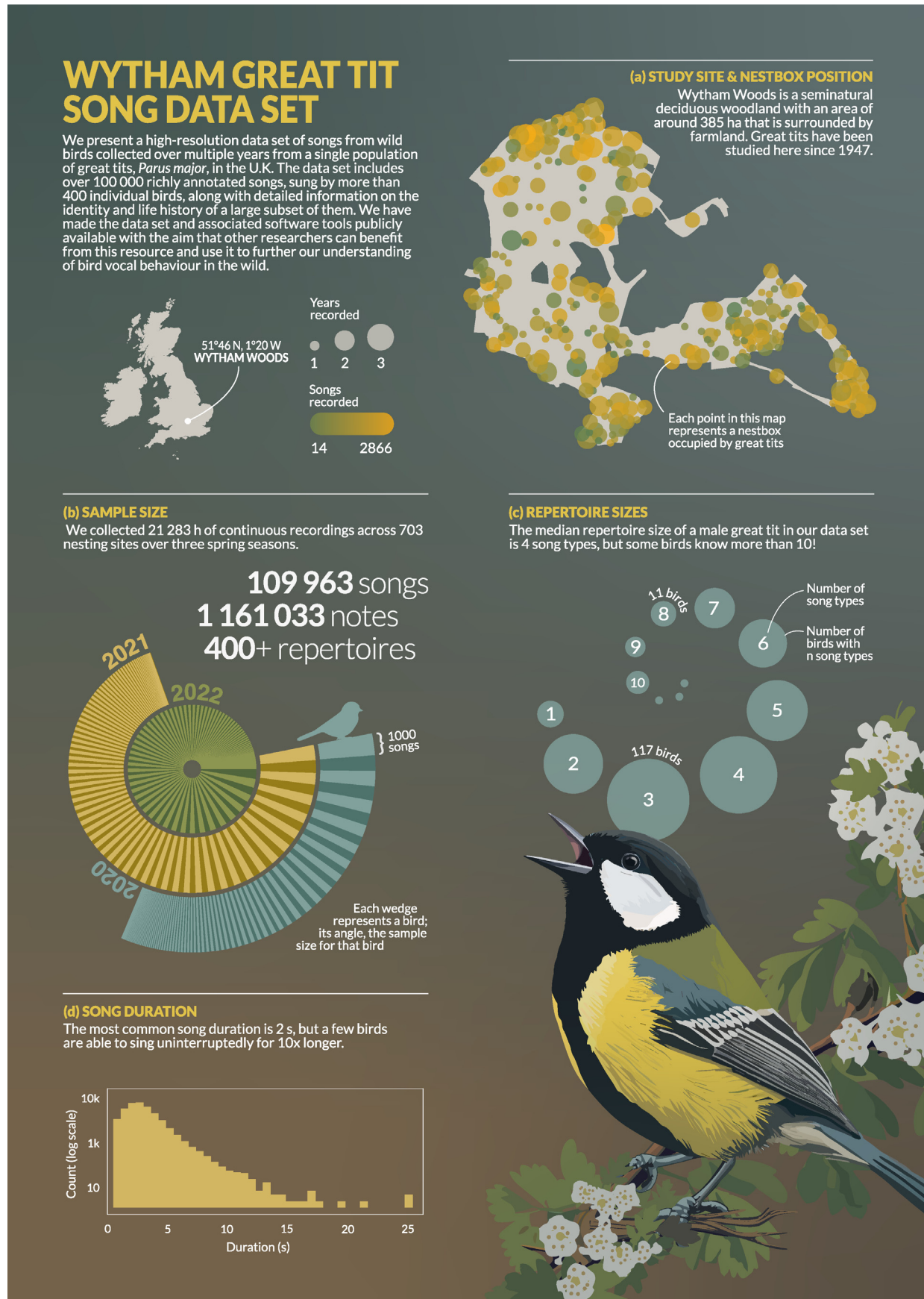


Figure 1. Visual summary of the Wytham Great Tit Song Data set. (a) Map of the study site and sample locations. (b) Total sample sizes for each bird and year. (c) Distribution of repertoire sizes. (d) Distribution of song lengths. Note that the number of individual birds is given as 454 but is not known exactly.

Ethical Note

All work involving birds was subject to review by the University of Oxford, Department of Zoology, Animal Welfare and Ethical Review Board (approval number: APA/1/5/ZOO/NASPA/Sheldon/Tit-BreedingEcology). Data collection adhered to local guidelines for the use of animals in research and all birds were caught, tagged and ringed by BTO licence holders.

Recording Equipment and Schedule

We used 60 (30 in 2020) AudioMoth recorders (Hill et al., 2019), which were housed in waterproof, custom-built enclosures. Recording began approximately 1 h before sunrise (0536-0400 UTC during the recording period) and consisted of seven consecutive 60 min recordings with a sample rate of 48 kHz, and a depth of 16-bit. To sample as many birds as possible, we left each recorder in the same location for at least 3 consecutive days before moving it to a different nestbox. We relocated 20 recorders (10 in 2020) every day throughout the recording period.

DATA PROCESSING AND ANNOTATION

We processed and annotated the recordings using custom software and scripts written in Python 3 (van Rossum, 1995), using the open-source package *pykanto* (Merino Recalde, 2023b). These are available from github.com/nilomr/great-tit-hits-setup (Merino Recalde, 2023a). Fig. 2c shows a graphic illustration of the process. See also the Appendix for a note on the terminology used for different parts of the songs.

Song Segmentation

We inspected spectrograms for each raw recording and selected songs based on a simple criterion: that its notes were clearly distinct from background noise and other bird vocalizations. We chose entire songs where it was possible; where it was not, we selected the longest contiguous segment possible. This process was carried out manually using the open-source software Sonic Visualiser (Cannam et al., 2010) by drawing boxes bounding songs in the time and frequency domains.

Assigning Song Bouts to Individuals

Due to the automated recording process, there is a possibility that some of the recorded songs near a particular nestbox may not originate from the focal bird. To minimize the chance of false positives, we discarded recordings with more than one vocalizing bird if one was not distinctly louder than the rest during the segmentation process. Additionally, we discarded all songs with a maximum amplitude below -16 dB, calculated as $20 \log_{10} \left(\frac{A}{A_0} \right)$, with $A = 5000$ and $A_0 = 32\,767$ (the maximum value for 16-bit digital audio). This specific threshold was derived from observations indicating that when simultaneous recordings captured neighbouring birds, an amplitude cutoff greater than 4000 consistently differentiated the focal bird from its closest neighbours. Note that these are not calibrated values and are, therefore, relative to the recording equipment and settings we used, as well as other factors like sound directionality and vegetation cover.

Spectrogramming

For most operations beyond this point, we used normalized, band-passed and log-scaled mel spectrogram representations of each of the songs (sampling rate = 22 050, window length = 1024,

hop length = 128, mel bins = 224; see the repository [nilomr/great-tit-hits-setup](https://github.com/nilomr/great-tit-hits-setup) for full details on the process).

Note Segmentation

We segmented the resulting song selections into their constituent notes using a custom dynamic threshold algorithm implemented in *pykanto* (Merino Recalde, 2023b), based on the work of Sainburg et al. (2019). Briefly, the algorithm finds minima in the spectral envelope of a spectrogram, which are considered silences; if the length of the signal between these minima exceeds a maximum note duration, a new local minimum is defined that divides the signal into two shorter segments. This is repeated until multiple notes are defined or there are no local minima below a maximum amplitude threshold. Then, segments below a minimum note duration threshold are discarded. To make the algorithm more robust to noise, the spectrogram is subject to morphological transformations and de-echoing before amplitude information is extracted. The de-echoing algorithm implemented in *pykanto* is based on that in Luscinia (Lachlan, 2016), and works by subtracting a delayed version of the spectrogram from itself. We determined minimum and maximum note length ranges by manually segmenting a small, random subset of songs ($N = 30$).

Note that the automated segmentation process is susceptible to various factors that can influence its accuracy. These include background noise, significant variation in amplitude between notes, attenuation caused by vegetation, changes in the direction of sound production and even variations in performance where some notes may be much quieter. As a result, the algorithm may fail to detect or incorrectly delimit certain notes. Despite this, we estimate that approximately 96% of the notes are correctly segmented (0.037 error rate based on a random subset of $N = 1048$ notes that were checked manually). Still, depending on specific goals, we recommend manual verification of note segmentation if complete accuracy is crucial.

Song Type Annotation

We annotated each song type in the data set using a semi-supervised approach implemented in *pykanto*. The process involved several steps to ensure accurate classification. First, we generated average unit spectrograms for each song by taking the mean of the centred and padded spectrograms of its units or notes, which provided a concise representation of the temporal and spectral characteristics of the syllable within it. Next, we performed nonlinear dimensionality reduction using UMAP (McInnes et al., 2018) and a cluster search using HDBSCAN (McInnes et al., 2017) for each bird in the data set. See Sainburg and Hedley (2020) and Thomas et al. (2021) for similar approaches. This strategy, while useful, often leads to spurious outcomes. For instance, it may separate renditions of the same song type if variation in performance or background noise exists, or if certain song elements are sometimes attenuated. Such variation could be misinterpreted as distinct song types, leading to an overestimation of repertoire size. To address this, we used the interactive app in *pykanto* to review and split or combine clusters as necessary for each bird. It is worth mentioning that this process would be significantly more challenging in species with highly variable songs: our approach benefited from the great tits' relatively limited repertoires (one to fewer than 15 song types in our population) and their tendency to produce stable and stereotyped songs.

Calculating Song Embeddings

Comparing animal vocalizations poses a significant challenge for researchers. Traditionally, two approaches have been used: visual comparisons of spectrograms and, more recently, measurement of

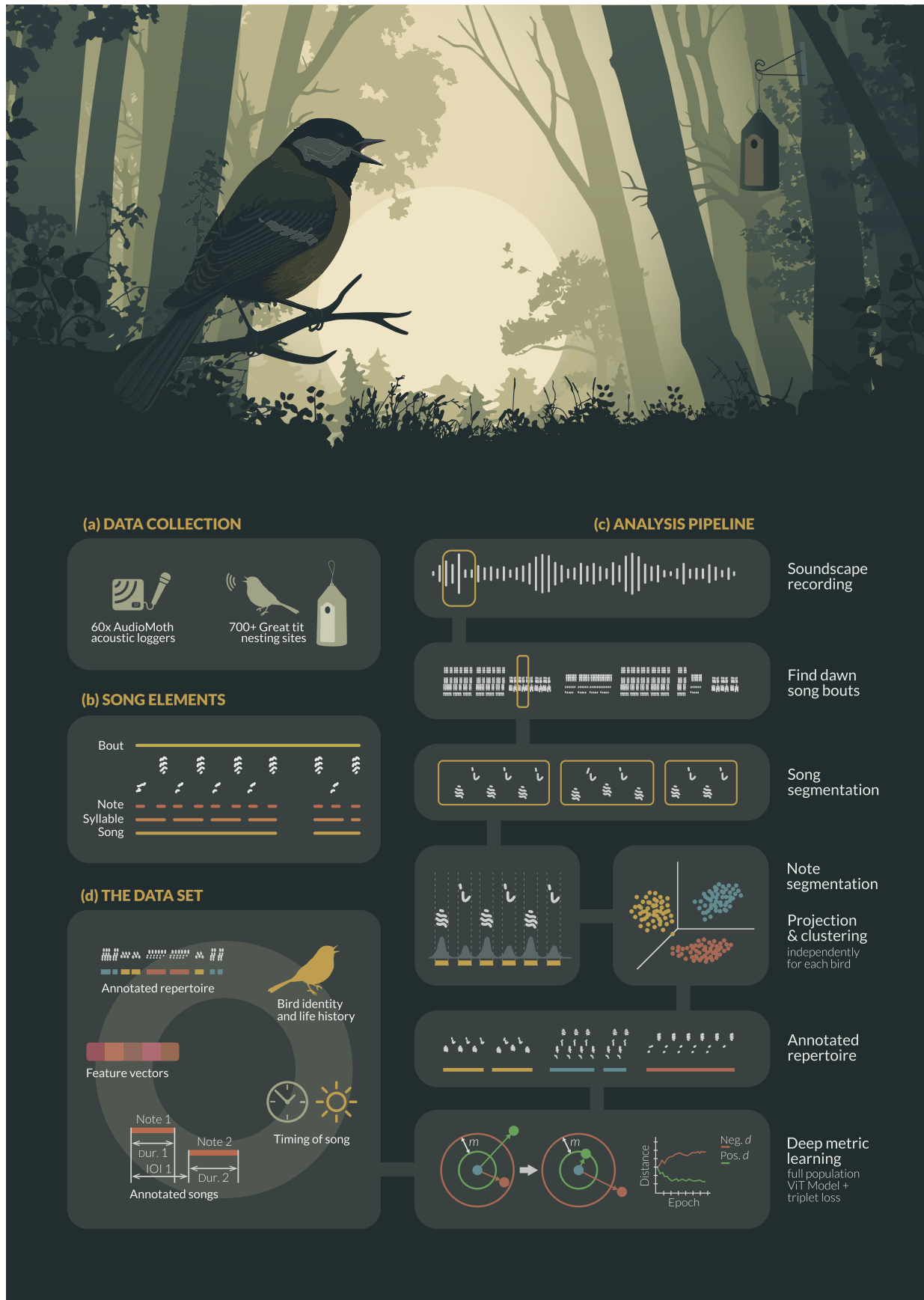


Figure 2. A brief visual summary of the data collection and analysis pipeline used to prepare the Wytham Great Tit Song Data set. (a) Data collection in the field. (b) The terminology used to describe the various hierarchical levels at which we can describe the great tit's singing. (c) Computational pipeline. (d) Main outputs included as part of the data set.

handpicked acoustic features (Goffinet et al., 2021). However, these methods have limitations when dealing with noise, variations in performance and changes in syntax (where compositional syntax is not relevant). For instance, if a song with the sequence ‘tea-cher, teacher’ is recorded as ‘cher-tea, cher’, it might be wrongly perceived as highly dissimilar, despite being the same song (see Stowell, 2021; Zandberg et al., 2022) for a good overview of these issues). Additionally, these methods often fail to capture high-level features such as the syntactic relationships between notes and other complex spectrotemporal characteristics that cannot be easily characterized by an orthogonal combination of simple acoustic features.

Unfortunately, we cannot rely on the birds’ perceptual judgments due to the lack of hard to obtain experimental data (although recent studies, such as Morfi et al., 2021; Zandberg et al., 2022, have explored this avenue). This can be an issue where the focus of research is behavioural interactions or the social functions of song. At the same time, for monitoring or individual identification purposes, fully mimicking the bird’s perceptual space may not be ideal: the performance of metric learning or classification algorithms trained for narrow purposes can surpass the organism’s abilities, as exemplified by facial recognition in humans (Lu & Tang, 2014). Here, our goal was to define a similarity space based on the inherent variation in the data and the only categorical labels that we know are perceptually and behaviourally significant: song types sung by individual birds. Given that great tits can recognize each other based on their vocalizations (Lind et al., 1996), we aimed to define a similarity space that facilitates similarity-based research and captures some of the song characteristics that birds themselves might attend to when distinguishing individuals. To do this, we took advantage of recent advances in the fields of deep learning and computer vision and used a data-driven approach. Below is a simple narrative description of the process. For further details, see the dedicated repository [nilomr/open-metric-learning](https://github.com/nilomr/open-metric-learning) and the OML library (Shabanov, 2023).

Metric learning with a vision transformer

Rather than focusing on classification, we aimed to develop semantically meaningful embeddings. To achieve this, we used a Vision Transformer (ViT) model as a feature extractor in a (Euclidean) metric learning task. These models, inspired by the success of transformers in natural language processing applications, process images by splitting them into patches, treating them as tokens similar to words in a natural language (Dosovitskiy et al., 2021; Raghu et al., 2022). In this case, we used the ViT-S/16 architecture (21.7 M parameters), pretrained on ImageNet using the DINO method (self-distillation with no labels; Caron et al., 2021).

Model training

During the training phase, we fine-tuned the ViT model using the great tit song data set. To optimize the performance of the model, we used Triplet loss, a loss function that ensures that the projection of a positive sample, which belongs to the same class as the anchor point, is closer to the anchor’s projection than that of a negative sample, which belongs to a different class, by at least a specified margin (Hermans et al., 2017; Hoffer & Ailon, 2018). This loss function enables embedding points of the same class to form clusters without collapsing into a single point, which allows us to also explore differences within song types. While training the model we mined hard triplets, where the negative sample is closer to the anchor than the positive, and used the Adam optimizer with a fixed learning rate of 1×10^{-5} .

Handling data imbalance and batch generation

The distribution of song sample sizes per individual in the great tit data set approximately follows a power law, resulting in a

significant data imbalance. Although the use of triplet loss already addresses this issue to some extent (Thakur et al., 2019), we adopt a random subsampling strategy where classes with more than 100 samples are reduced to 100 for computational efficiency, classes with fewer than 15 samples are excluded to allow a large enough query/gallery split for validation, and we ensure fair representation during training using a balanced sampler (Hermans et al., 2017). Our batch generation strategy involves uniformly sampling P song types without replacement and sampling K spectrograms for each song type, with replication as necessary. This guarantees that all labels are selected at least once in each epoch.

Train-time data augmentation

To enhance model robustness and prevent overfitting, we apply various train-time data augmentation techniques (Mumuni & Mumuni, 2022; Perez & Wang, 2017; Shorten & Khoshgofaar, 2019). These include random cropping in the time domain, dropping out parts of the spectrogram, adding Gaussian and multiplicative noise, equalization, sharpening, changes to brightness and contrast, blurring, and slight shifting in both time and frequency domains. The latter augmentations are applied within the typical variation in performance observed in the great tit vocalizations.

Results

Our trained model shows very good performance, achieving a mean average precision at 5 (mAP@5) of 0.98 and a cumulative matching characteristic at 1 (CMC@1) of 0.98. This indicates that in approximately 98% of the queries made to the similarity space, the returned candidate song type by a bird is the correct one. Errors primarily stemmed from instances where songs of the same type sung by the same bird appeared more than once in the data set, which happened if a bird survived to the next year. Given that the model was trained on almost 2000 classes, this means that there is enough individual information contained in each song type to distinguish between birds with high confidence, which has important implications for both the study of individuality and population monitoring. See Fig. 3 for a visual representation of the embedding space and nearest-neighbour queries.

DATA RECORDS AND DESCRIPTION

Table A1 contains a summary of the files included with the data set. Detailed data documentation, including variable descriptions, can be found online at [nilomr.github.io/great-tit-hits](https://github.com/nilomr/great-tit-hits).

The data set provides a comprehensive view of the populations’ natural dawn singing behaviour over three spring seasons. It documents changes in individual performance, the appearance and disappearance of birds, and with them their songs, and highlights just how much behavioural variation there is along every dimension of what could at first seem a relatively simple trait. Table 1 presents some simple summary statistics and Fig. 1 provides a visual overview of the data set.

Even though most birds in the data set are 1- or 2-year-olds recorded within a single year (which can be attributed to high turnover rates in the population given low annual survival), the data set includes valuable data on much older individuals, including a 7-year-old. Among the recorded birds, some display metronome-like regularity in their performance, while others have highly variable or unusual songs, due to learning from allospecific vocalizations, or even issues with their vocal apparatus. You can find some interactive examples at [nilomr.github.io/great-tit-hits](https://github.com/nilomr/great-tit-hits). The longest song recorded is approximately 20 times longer than the shortest song (and, coincidentally, was sung by one of the largest great tits ever recorded in the Wytham population). The median number of songs per song type and per bird in the data set

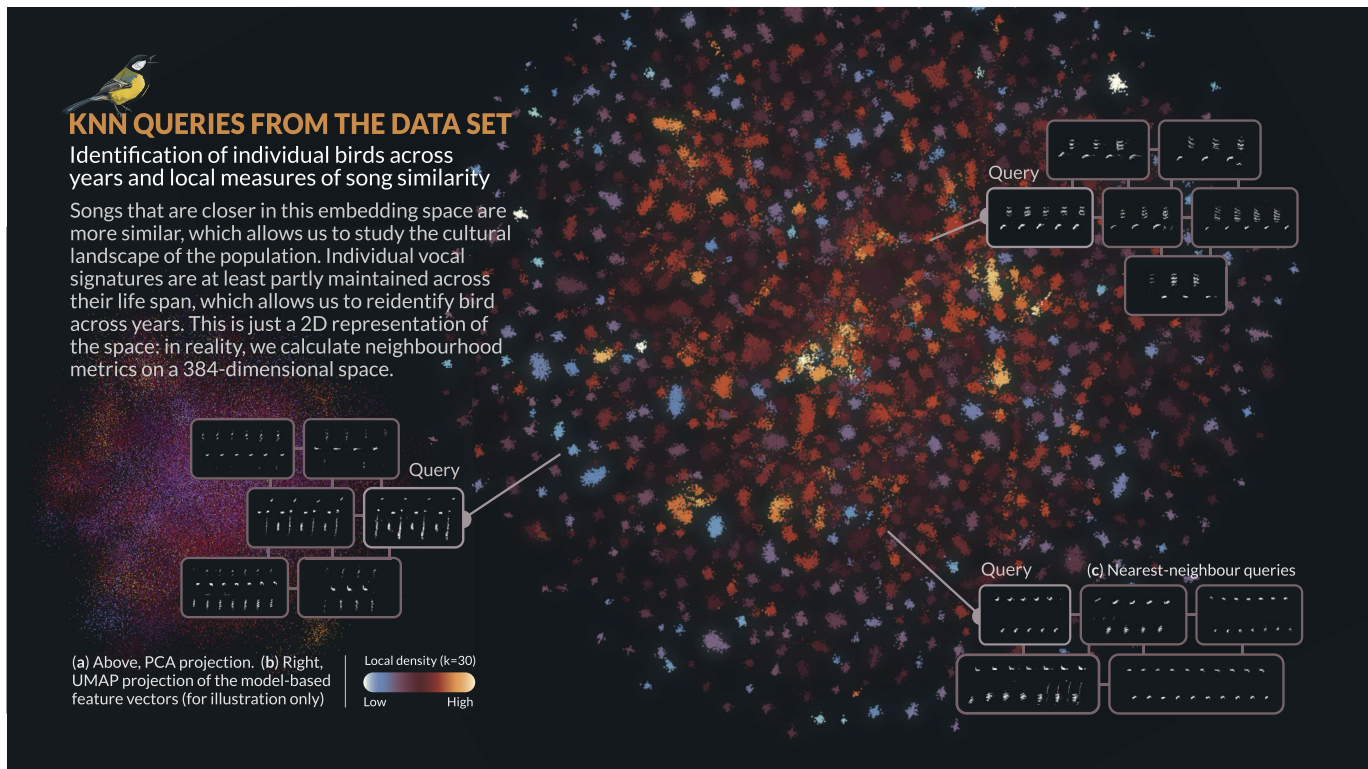


Figure 3. Measuring similarity is a very hard problem, in large part because there is often no objective way to compare the performance of different methods. Here, we took a data-based approach by training a Vision Transformer (ViT) model as a feature extractor in a Euclidean metric learning task. The resulting embedding space allows us to judge whether two songs are very similar, and to reidentify birds. (a) PCA projection of the feature vectors: two orthogonal linear components do not capture much of the high-level distinguishing features. (b) This figure shows a UMAP projection of the 384-dimensional vectors for each song in the data set into 2D, which leads to an arbitrary but useful visualization where tight clusters of points correspond to song types in the repertoire of individual birds. They are coloured by how densely occupied that region of space is in the high-dimensional space, based on $k = 30$ neighbours from other song types. (c) A k -nearest-neighbour search returns the closest matches for a query vector (highlighted).

is 31, with a significant number of birds having a much larger count, reaching into the thousands. Additionally, the median repertoire size per bird is four distinct song types, although some birds performed as many as 13 distinct song types.

Known Biases and Problems

Working with third-party data sets can be challenging, perhaps particularly so in the study of behaviour in natural populations. The familiarity that fieldworkers inevitably develop with the study system and the data is difficult to replace, and, as a result, there is a risk of unintentionally overlooking important sources of bias and variability. We have compiled a list of some key considerations,

Table 1
Brief description of the data set and sample sizes

Description	Value
Number of segmented notes	1 161 033
Number of songs	109 963
Mean repertoire size	4.24
SD repertoire size	1.98
Median repertoire size	4
Number of unique classes	1 930
Mean class size	56.98
SD class size	68.35
Median class size	31
Number of nest sites recorded	706
Number of nest sites with data	454
Number of unique birds with known ID	242
Number of times each bird was recorded	192 (1 year), 42 (2 years), 8 (3 years)

which, while not exhaustive, can serve as a starting point for identifying and addressing biases when testing hypotheses, estimating parameters or evaluating findings from the data. These issues can be broadly classified into two groups: those around bird and song type labelling, which can be partially addressed, and those that are inherent in the data or how they were collected.

Individual and song identification

One factor that can be partially addressed is that the birds recorded in our data set are not a random subset of the population; they are those that establish territories and begin the breeding process. In turn, birds that are subsequently identified are more likely to be those whose chicks hatch and survive for at least 6 days, when the first identification attempt is made. This may skew the distribution of certain behaviours within the data set or lead to endogenous selection bias (Elwert & Winship, 2014). One way to quantify the extent to which the subset of identified birds is representative of the entire breeding population would be to compare the distribution of the trait of interest in both groups. See, for example, Kidd et al. (2015), who found that females in nests that fail early in our population are more likely to be immigrant birds breeding in poor-quality areas.

Another issue to consider is that birds may attempt to breed again in the same nestbox or elsewhere after a failed attempt. This, coupled with a failure to identify the male associated with those attempts, means that it is conceivable (although likely very rare) that songs from the same bird could appear in the same year twice, leading to pseudoreplication. Similarly, unidentified birds present in the data set for multiple years could contribute to this problem.

One potential way to address these issues is by using song embeddings for identification based on similarity and assigning dummy IDs to birds believed to be the same individual. At least, this should be modelled to assess the sensitivity of any results to varying degrees of pseudoreplication from this source.

Finally, a few songs might have been mislabelled before model training, as it is not feasible to manually check such a large data set. However, the model-based embeddings can help identify any mislabelled songs: they will be clear outliers within their respective classes, thanks to the relatively discrete nature of great tit repertoires.

Unequal samples, songs and calls, and female song

As is common in many complex systems, the interaction of the many processes involved in both song production and sampling results in a heavy-tailed frequency distribution of sample sizes. This variation stems from various sources, including characteristics inherent to the study system, such as individual differences in singing activity and temporal fluctuations throughout the spring season. The sampling process introduces further variation, through factors like equipment malfunctions causing small gaps in the data, variation in recording dates relative to peak activity, and the impact of rain and hail on singing activity and recording quality. We cannot

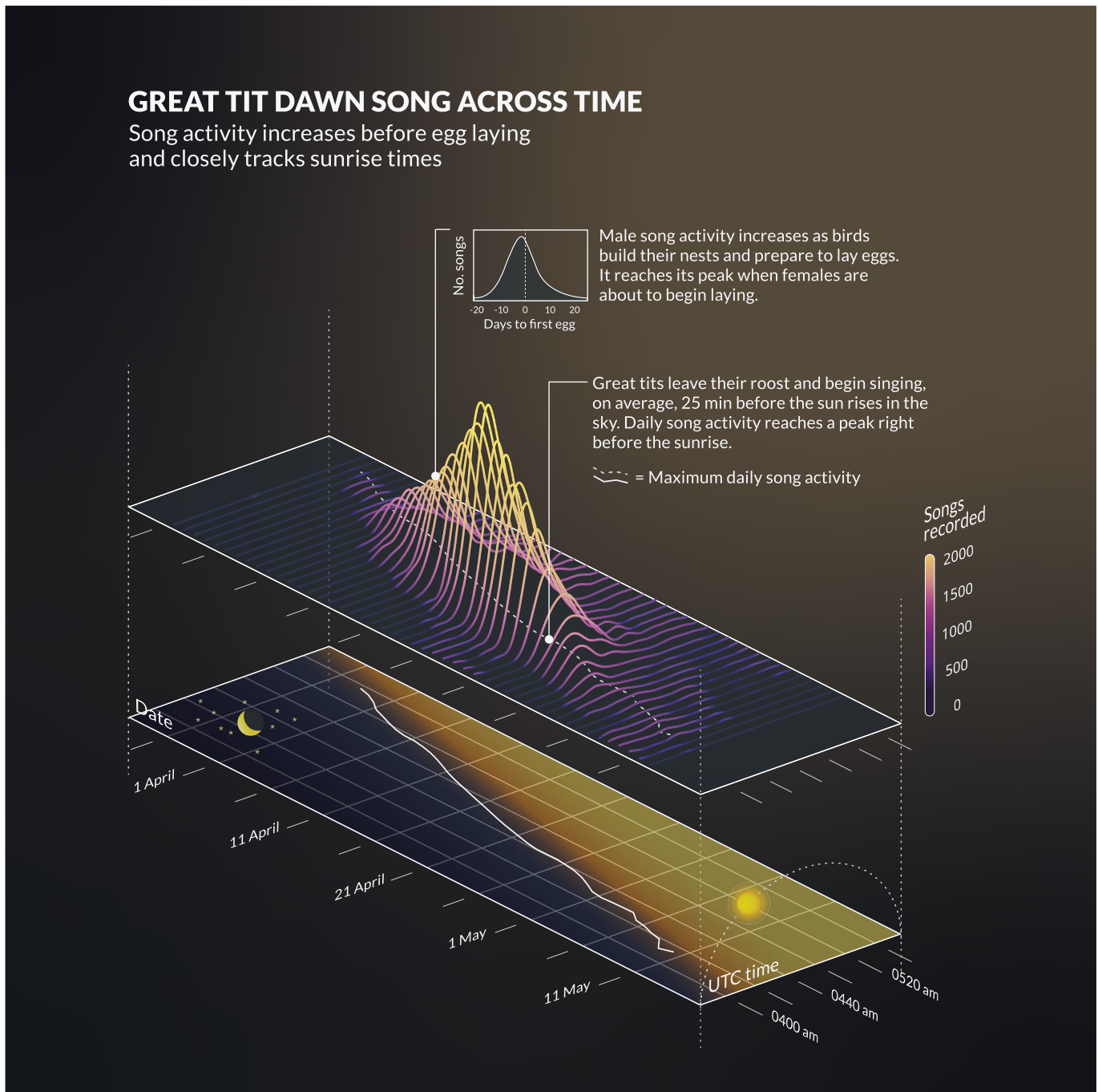


Figure 4. Days get longer as the spring progresses and male great tits track the advancing sunrise times with great precision, so that they always begin singing, on average, 25 min before the morning breaks. This figure also shows (z-axis) how song activity peaks alongside egg laying: males sing the most in the morning right before their partner lays the first egg.

assume these processes to be completely independent of each other. Therefore, when analysing song output or repertoire size, it is important to explicitly specify the assumed causal relationship between factors such as individual characteristics, sampling probability and the outcome measure.

Another important aspect to consider is that, while we have said that the data set consists of songs, the demarcation between songs and calls is not entirely straightforward. Some vocalizations that would typically be classified as calls, due to their acoustically simpler, shorter and possibly more stereotyped nature, are actually used as part of the dawn vocal behaviour. These vocalizations are repeated in a manner that creates an impression of functional equivalence to songs. While we have followed criteria similar to other studies (Baker et al., 1986; Fayet et al., 2014; Krebs et al., 1978; Rivera-Gutierrez, Matthysen et al., 2010) to maintain consistency, we believe that this phenomenon warrants further attention. These calls were not segmented and thus are not included in the data set, but we are happy to provide soundscape recordings to anyone interested in exploring this aspect further.

Finally, although female song in birds has received relatively little historical attention (see Langmore, 2020; Odom & Benedict, 2018; Riebel et al., 2005 for further discussion), female great tits also sing (see a brief treatment in Gompertz, 1961; Hinde, 1952). The vast majority of songs in the data set belong to the dawn song, a behaviour exclusively performed by the male prior to the female leaving the nest (a pattern observed in blue tits, *Cyanistes caeruleus*, as well, as documented by Sierro et al., 2022). Females, on the other hand, vocalize within the nest, but these vocalizations differ from songs (Gorissen & Eens, 2004, 2005) and were not typically detectable by our recording devices. Nevertheless, Hinde (1952) suggested that in the absence of males, females may be more inclined to engage in territorial behaviour that involves singing rather than just producing calls. If that is the case, it is possible that our data set contains some isolated instances of female song.

USES AND SUGGESTIONS

The data set we are presenting contains detailed information about the vocal behaviour and life of wild birds, providing valuable opportunities for investigating a wide range of research questions. In this section, we suggest several research areas that can be explored using this data set and provide references to relevant studies in the literature.

Behavioural Repeatability and Stability Across Multiple Scales

Researchers can use the data set to examine the repeatability and stability of song production and song characteristics across different temporal and spatial scales. This includes studying consistency in vocal behaviour within individuals over time and across different contexts, and its links to age (Rivera-Gutierrez et al., 2012; Zipple et al., 2019) and reproductive fitness (Sierro et al., 2023).

Links Between Vocal Performance or Diversity and Reproductive Success

Our data can be used to explore the relationships between vocal performance metrics, such as song complexity or vocal diversity, and individual breeding success on a data set that is much larger than what is typical in the field (Beecher et al., 2020; Crates et al., 2021; Hiebert et al., 1989; Hutfluss et al., 2022; McGregor et al., 1981).

Spatial and Temporal Properties of Acoustic Communities

The data set enables investigations into the spatial properties of acoustic communities, including the distribution of singing individuals within a given habitat and across time. This can provide valuable insights into the spatial dynamics of communication networks and acoustic interaction among neighbour birds.

Timing and Volume of Song Production

Researchers can use the data set to analyse the temporal patterns and timing of song production in great tits. This might involve studying diurnal variation, seasonal trends and the influence of environmental factors on the timing and abundance of vocal behaviour. As an example, Fig. 4 provides an overview of key temporal shifts in dawn singing behaviour: male birds sing more during the fertile period of the female, and their activity closely tracks advancing sunrise times.

The Syntactic Organization of Song Production

The data set captures song activity over entire dawn song periods, across days, and even years for many individuals. This would allow researchers to investigate the set of rules that govern the arrangement of song elements and transitions within the vocal repertoire of wild great tits, in terms of short- and long-distance dependencies and other properties of their sequential dynamics (Hedley et al., 2018; Lachlan et al., 2010; Sainburg et al., 2019; Searcy et al., 2022).

Song Learning in the Wild

While this data set does not directly provide evidence of song learning, researchers can use song similarity and proximity in time and space to infer cultural transmission processes. This allows for the exploration of the influence of spatial and social factors on song learning (James et al., 2020; Lachlan & Slater, 2003; Nelson & Poesel, 2014; Peters & Nowicki, 2017; Wheelwright et al., 2008).

Conclusion

With over 1 100 000 annotated notes and acoustic units from more than 100 000 songs, collected over three spring seasons, we hope that this data set will offer valuable insights into bird vocal behaviour and song culture. The data set is enriched with detailed metadata such as note onset and offset times, song type labels and embeddings derived from a deep metric learning model, as well as identity and life history information for the birds, which makes it useful for a wide range of research purposes. By sharing this comprehensive data set, we also aim to help promote data sharing and scientific collaboration.

Author Contributions

L.P., A.V., A.E. and N.M.R. collected the data. N.M.R. created the software and pipeline to plan fieldwork and analyse the data, annotated the data set with help from A.E., built the website, documentation and visualizations, and wrote the original draft. B.C.S. and E.F.C. provided feedback and supervision throughout the research, and all authors contributed critically to the manuscript.

Data Availability

The complete Wytham Great Tit Song Data Set and its metadata are available at [10.17605/OSF.IO/N8AC9](https://doi.org/10.17605/OSF.IO/N8AC9) (Merino Recalde, 2023c). The code to train the deep metric learning model can be found at [nilomr/open-metric-learning](https://github.com/nilomr/open-metric-learning). See the data set website for documentation and more information. All input and output data files use open data formats and are under a CC-BY-4.0 licence. The scripts and software used to create this data set are available under the MIT licence from GitHub [nilomr/great-tit-hits-setup](https://github.com/nilomr/great-tit-hits-setup) and archived at Zenodo (Merino Recalde, 2023a). This means that you may share, copy and redistribute the data and code in any medium or format; adapt, remix, transform and build upon it for any purpose, even commercially, as long as you give appropriate credit, provide a link to the licence and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the authors endorse you or your use.

Declaration of Interest

The authors declare no conflict of interest.

Acknowledgments

We thank all those who have contributed to the long-term nestbox study in Wytham Woods and the collection of associated data. This work was supported by a Clarendon-Mary Frances Wagley Graduate Scholarship and an Edward Grey Institute scholarship to Nilo Merino Recalde and made use of the University of Oxford Advanced Research Computing facility (Richards, 2015).

References

- Aplin, L. M. (2019). Culture and cultural evolution in birds: A review of the evidence. *Animal Behaviour*, *147*, 179–187. <https://doi.org/10.1016/j.anbehav.2018.05.001>
- Baker, M. C., Bjerke, T. K., Lampe, H., & Espmark, Y. (1986). Sexual response of female great tits to variation in size of males' song repertoires. *American Naturalist*, *128*(4), 491–498. <https://doi.org/10.1086/284582>
- Beecher, M. D., Akçay, Ç., & Campbell, S. E. (2020). Birdsong learning is mutually beneficial for tutee and tutor in song sparrows. *Animal Behaviour*, *166*, 281–288. <https://doi.org/10.1016/j.anbehav.2020.05.015>
- Behr, O., & von Helversen, O. (2004). Bat serenades complex courtship songs of the sac-winged bat (*Saccopteryx bilineata*). *Behavioral Ecology and Sociobiology*, *56*(2), 106–115. <https://doi.org/10.1007/s00265-004-0768-7>
- Bell, M. B. V., Radford, A. N., Smith, R. A., Thompson, A. M., & Ridley, A. R. (2010). Bargaining babblers: Vocal negotiation of cooperative behaviour in a social bird. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1698), 3223–3228. <https://doi.org/10.1098/rspb.2010.0643>
- Cannam, C., Landone, C., & Sandler, M. (2010). Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. In *Proceedings of the International Conference on Multimedia – MM '10* (p. 1467). Firenze, Italy: ACM Press. <https://doi.org/10.1145/1873951.1874248>
- Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). *Emerging properties in self-supervised vision transformers*. <https://doi.org/10.48550/arXiv.2104.14294>. arXiv.
- Crates, R., Langmore, N., Ranjard, L., Stojanovic, D., Rayner, L., Ingwersen, D., & Heinsohn, R. (2021). Loss of vocal culture and fitness costs in a critically endangered songbird. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1947), Article 20210225. <https://doi.org/10.1098/rspb.2021.0225>
- Davenport, M. H., & Jarvis, E. D. (2023). Birdsong neuroscience and the evolutionary substrates of learned vocalization. *Trends in Neurosciences*, *46*(2), 97–99. <https://doi.org/10.1016/j.tins.2022.11.005>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissensborn, D., Zhai, X., Unterthiner, T., & Houlsby, N. (2021). *An image is worth 16x16 words: Transformers for image recognition at scale*. <https://doi.org/10.48550/arXiv.2010.11929>. arXiv.
- Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, *40*, 31–53. <https://doi.org/10.1146/annurev-soc-071913-043455>
- Engesser, S., & Manser, M. B. (2022). Collective close calling mediates group cohesion in foraging meerkats via spatially determined differences in call rates. *Animal Behaviour*, *185*, 73–82. <https://doi.org/10.1016/j.anbehav.2021.12.014>
- Fayet, A. L., Tobias, J. A., Hintzen, R. E., & Seddon, N. (2014). Immigration and dispersal are key determinants of cultural diversity in a songbird population. *Behavioral Ecology*, *25*(4), 744–753. <https://doi.org/10.1093/beheco/aru047>
- Gerhardt, H. C. (1991). Female mate choice in treefrogs: Static and dynamic acoustic criteria. *Animal Behaviour*, *42*(4), 615–635. [https://doi.org/10.1016/S0003-3472\(05\)80245-3](https://doi.org/10.1016/S0003-3472(05)80245-3)
- Goffinet, J., Brudner, S., Mooney, R., & Pearson, J. (2021). Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires. *Elife*, *10*, Article e67855. <https://doi.org/10.7554/eLife.67855>
- Gompertz, T. (1961). The vocabulary of the great tit. *British Birds*, *54*(10–11), 369–394.
- Gorissen, L., & Eens, M. (2004). Interactive communication between male and female great tits (*Parus major*) during the dawn chorus. *Auk*, *121*(1), 184–191.
- Gorissen, L., & Eens, M. (2005). Complex female vocal behaviour of great and blue tits inside the nesting cavity. *Behaviour*, *142*(4), 489–506. <http://www.jstor.org/stable/4536256>
- Halfwerk, W., Bot, S., & Slabbekoorn, H. (2012). Male great tit song perch selection in response to noise-dependent female feedback. *Functional Ecology*, *26*(6), 1339–1347. <https://doi.org/10.1111/j.1365-2435.2012.02018.x>
- Hedley, R. W., Logue, D. M., Benedict, L., & Mennill, D. J. (2018). Assessing the similarity of song-type transitions among birds: Evidence for interspecies variation. *Animal Behaviour*, *140*, 161–170. <https://doi.org/10.1016/j.anbehav.2018.04.008>
- Hermans, A., Beyer, L., & Leibe, B. (2017). *In defense of the triplet loss for person re-identification*. <https://doi.org/10.48550/arXiv.1703.07737>. arXiv.
- Hersh, T. A., Ravignani, A., & Burchard, L. S. (2023). Robust rhythm reporting will advance ecological and evolutionary research. *Methods in Ecology and Evolution*. <https://doi.org/10.1111/2041-210X.14118>
- Hiebert, S. M., Stoddard, P. K., & Arcese, P. (1989). Repertoire size, territory acquisition and reproductive success in the song sparrow. *Animal Behaviour*, *37*, 266–273. [https://doi.org/10.1016/0003-3472\(89\)90115-2](https://doi.org/10.1016/0003-3472(89)90115-2)
- Hill, A. P., Prince, P., Snaddon, J. L., Doncaster, C. P., & Rogers, A. (2019). AudioMoth: A low-cost acoustic device for monitoring biodiversity and the environment. *HardwareX*, *6*, Article e00073. <https://doi.org/10.1016/j.ohx.2019.e00073>
- Hinde, R. A. (1952). The behaviour of the great tit (*Parus major*) and some other related species. *Behaviour*, (2), III–201 <http://www.jstor.org/stable/30039125>
- Hoffer, E., & Ailon, N. (2018). *Deep metric learning using Triplet network*. <https://doi.org/10.48550/arXiv.1412.6622>. arXiv.
- Hutflus, A., Bermúdez-Cuamatzin, E., Mouchet, A., Briffa, M., Slabbekoorn, H., & Dingemans, N. J. (2022). Male song stability shows cross-year repeatability but does not affect reproductive success in a wild passerine bird. *Journal of Animal Ecology*, *91*(7), 1507–1520. <https://doi.org/10.1111/1365-2656.13736>
- James, L. S., Davies, R., Mori, C., Wada, K., & Sakata, J. T. (2020). Manipulations of sensory experiences during development reveal mechanisms underlying vocal learning biases in zebra finches. *Developmental Neurobiology*, *80*(3–4), 132–146. <https://doi.org/10.1002/dneu.22754>
- Jenkins, G. B., Beckerman, A. P., Bellard, C., Benítez-López, A., Ellison, A. M., Foote, C. G., & Peres-Neto, P. R. (2023). Reproducibility in ecology and evolution: Minimum standards for data and code. *Ecology and Evolution*, *13*(5), Article e9961. <https://doi.org/10.1002/ece3.9961>
- Kacelnik, A., & Krebs, J. R. (1983). The dawn chorus in the great tit (*Parus major*): Proximate and ultimate causes. *Behaviour*, *83*(3–4), 287–308. <https://doi.org/10.1163/156853983X00200>
- Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, *61*, Article 101236. <https://doi.org/10.1016/j.ecoinf.2021.101236>
- Kidd, L. R., Sheldon, B. C., Simmonds, E. G., & Cole, E. F. (2015). Who escapes detection? Quantifying the causes and consequences of sampling biases in a long-term field study. *Journal of Animal Ecology*, *84*(6), 1520–1529. <https://doi.org/10.1111/1365-2656.12411>
- Kollmorgen, S., Hahnloser, R. H. R., & Mante, V. (2020). Nearest neighbours reveal fast and slow components of motor learning. *Nature*, *577*(7791), 526–530. <https://doi.org/10.1038/s41586-019-1892-x>
- Krebs, J. R. (1976). Habituation and song repertoires in the great tit. *Behavioral Ecology and Sociobiology*, *1*(2), 215–227. <https://doi.org/10.1007/BF00299198>
- Krebs, J., Ashcroft, R., & Webber, M. (1978). Song repertoires and territory defence in the great tit. *Nature*, *271*(5645), 539–542. <https://doi.org/10.1038/271539a0>
- Lachlan, R. F. (2016). *Luscinia: A bioacoustics analysis computer program*.
- Lachlan, R. F., Ratmann, O., & Nowicki, S. (2018). Cultural conformity generates extremely stable traditions in bird song. *Nature Communications*, *9*(1), 2417. <https://doi.org/10.1038/s41467-018-04728-1>
- Lachlan, R. F., & Slater, P. J. (2003). Song learning by chaffinches: How accurate, and from where? *Animal Behaviour*, *65*(5), 957–969. <https://doi.org/10.1006/anbe.2003.2091>
- Lachlan, R. F., Verhagen, L., Peters, S., & ten Cate, C. (2010). Are there species-universal categories in bird song phonology and syntax? A comparative study of chaffinches (*Fringilla coelebs*), zebra finches (*Taenopygia guttata*), and swamp sparrows (*Melospiza georgiana*). *Journal of Comparative Psychology*, *124*(1), 92–108. <https://doi.org/10.1037/a0016996>
- Lack, D. (1964). A long-term study of the great tit (*Parus major*). *Journal of Animal Ecology*, *33*, 159–173. <https://doi.org/10.2307/2437>
- Lambrechts, M. M., & Dhondt, A. A. (1990). A relationship between the composition and size of great tit song repertoires. *Animal Behaviour*, *39*(2), 213–218. [https://doi.org/10.1016/S0003-3472\(05\)80865-6](https://doi.org/10.1016/S0003-3472(05)80865-6)
- Langmore, N. E. (2020). Female birdsong. *Current Biology*, *30*(14), R789–R790. <https://doi.org/10.1016/j.cub.2020.05.042>
- Lehmann, J. A. M., & Seufert, T. (2017). The influence of background music on learning in the light of different theoretical perspectives and the role of

- working memory capacity. *Frontiers in Psychology*, 8(October), 1–11. <https://doi.org/10.3389/fpsyg.2017.01902>
- Lind, H., Dabelsteen, T., & McGregor, P. K. (1996). Female great tits can identify mates by song. *Animal Behaviour*, 52(4), 667–671. <https://doi.org/10.1006/anbe.1996.0211>
- Linhart, P., Osiejuk, T. S., Budka, M., Sálek, M., Špinka, M., Policht, R., & Blumstein, D. T. (2019). Measuring individual identity information in animal signals: Overview and performance of available identity metrics. *Methods in Ecology and Evolution*, 10(9), 1558–1570. <https://doi.org/10.1111/2041-210X.13238>
- Lu, C., & Tang, X. (2014). Surpassing human-level face verification performance on LFW with GaussianFace. <https://doi.org/10.48550/arXiv.1404.3840>. arXiv.
- Mace, R. (1987). The dawn chorus in the great tit *Parus major* is directly related to female fertility. *Nature*, 330(6150), 745–746. <https://doi.org/10.1038/330745a0>
- Marler, P. (1952). Variation in the song of the Chaffinch *Fringilla coelebs*. *Ibis*, 94(3), 458–472. <https://doi.org/10.1111/j.1474-919X.1952.tb01845.x>
- Marler, P., & Tamura, M. (1962). Song 'dialects' in three populations of white-crowned sparrows. *Condor*, 64(5), 368–377. <https://doi.org/10.2307/1365545>
- Marler, P., & Tamura, M. (1964). Culturally transmitted patterns of vocal behavior in sparrows. *Science*, 146(3650), 1483–1486. <https://doi.org/10.1126/science.146.3650.1483>
- McGregor, P. K., & Krebs, J. R. (1982). Song types in a population of great tits (*Parus major*): Their distribution, abundance and acquisition by individuals. *Behaviour*, 79(2/4), 126–152. <http://www.jstor.org/stable/4534157>.
- McGregor, P. K., & Krebs, J. R. (1989). Song learning in adult great tits (*Parus major*): Effects of neighbours. *Behaviour*, 108(1–2), 139–159. <https://doi.org/10.1163/156853989X00105>
- McGregor, P. K., Krebs, J. R., & Perrins, C. M. (1981). Song repertoires and lifetime reproductive success in the great tit (*Parus major*). *American Naturalist*, 118(2), 149–159.
- Mcgregor, P. K., Krebs, J. R., & Ratcliffe, L. M. (1983). The reaction of great tits (*Parus major*) to playback of degraded and undegraded songs: The effect of familiarity with the stimulus song type. *Auk*, 100(4), 898–906.
- McInnes, L., Healy, J., & Astels, S. (2017). HdbSCAN: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11), 205. <https://doi.org/10.21105/joss.00205>
- McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. *Journal of Open Source Software*, 3(29), 861. <https://arxiv.org/abs/1802.03426>.
- Merino Recalde, N. (2023a). Nilomr/great-tit-hits-setup: V0.1.0. Zenodo. <https://doi.org/10.5281/zenodo.8092397>.
- Merino Recalde, N. (2023b). Pykanto: A python library to accelerate research on wild bird song. *Methods in Ecology and Evolution*, 00, 1–9. <https://doi.org/10.1111/2041-210X.14155>
- Merino Recalde, N. (2023c). Wytham great tit song dataset. The open science framework. <https://doi.org/10.17605/OSF.IO/N8AC9>.
- Mets, D. G., & Brainard, M. S. (2019). Learning is enhanced by tailoring instruction to individual genetic differences. *Elife*, 8, 1–15. <https://doi.org/10.7554/eLife.47216>
- Morfi, V., Lachlan, R. F., & Stowell, D. (2021). Deep perceptual embeddings for unlabelled animal sound events. *Journal of the Acoustical Society of America*, 150(1), 2–11. <https://doi.org/10.1121/1.510005475>
- Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, 16, Article 100258. <https://doi.org/10.1016/j.array.2022.100258>
- Nelson, D. A., & Poesel, A. (2014). Tutor choice and imitation accuracy during song learning in a wild population of the Puget Sound white-crowned sparrow. *Behavioral Ecology and Sociobiology*, 68(11), 1741–1752. <https://doi.org/10.1007/s00265-014-1782-z>
- Odom, K. J., & Benedict, L. (2018). A call to document female bird songs: Applications for diverse fields. *Auk*, 135(2), 314–325. <https://doi.org/10.1642/AUK-17-183.1>
- Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. <https://doi.org/10.48550/arXiv.1712.04621>. arXiv.
- Perrins, C. M. (1965). Population fluctuations and clutch-size in the great tit, *Parus major* L. *Journal of Animal Ecology*, 34(3), 601–647.
- Peters, S., & Nowicki, S. (2017). Overproduction and attrition: The fates of songs memorized during song learning in songbirds. *Animal Behaviour*, 124, 255–261. <https://doi.org/10.1016/j.anbehav.2016.09.019>
- Pitcher, B. J., Harcourt, R. G., & Charrier, I. (2010). Rapid onset of maternal vocal recognition in a colonially breeding mammal, the Australian sea lion. *PLoS One*, 5(8), Article e12195. <https://doi.org/10.1371/journal.pone.0012195>
- Powers, S. M., & Hampton, S. E. (2019). Open science, reproducibility, and transparency in ecology. *Ecological Applications*, 29(1), Article e01822. <https://doi.org/10.1002/eap.1822>
- Radford, A. N., & Ridley, A. R. (2007). Individuals in foraging groups may use vocal cues when assessing their need for anti-predator vigilance. *Biology Letters*, 3(3), 249–252. <https://doi.org/10.1098/rsbl.2007.0110>
- Raghu, M., Unterthiner, T., Kornblith, S., Zhang, C., & Dosovitskiy, A. (2022). Do vision transformers see like convolutional neural networks?. <https://doi.org/10.48550/arXiv.2108.08810>. arXiv.
- Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. *Science*, 331(6018), 703–705. <https://doi.org/10.1126/science.1197962>
- Richards, A. (2015). University of oxford advanced research computing [Manual]. <https://doi.org/10.5281/zenodo.22558>.
- Riebel, K., Hall, M., & Langmore, N. (2005). Female songbirds still struggling to be heard. *Trends in Ecology & Evolution*, 20, 419–420. <https://doi.org/10.1016/j.tree.2005.04.024>
- Riebel, K., Lachlan, R. F., & Slater, P. J. (2015). Learning and cultural transmission in chaffinch song. *Advances in the Study of Behavior*, 47, 181–227. <https://doi.org/10.1016/bs.asb.2015.01.001>
- Ritschard, M., van Oers, K., Naguib, M., & Brumm, H. (2012). Song amplitude of rival males modulates the territorial behaviour of great tits during the fertile period of their mates. *Ethology*, 118(2), 197–202. <https://doi.org/10.1111/j.1439-0310.2011.01999.x>
- Rivera-Gutierrez, H. F., Matthysen, E., Adriaensen, F., & Slabbekoorn, H. (2010). Repertoire sharing and song similarity between great tit males decline with distance between forest fragments. *Ethology*, 116(10), 951–960. <https://doi.org/10.1111/j.1439-0310.2010.01813.x>
- Rivera-Gutierrez, H. F., Pinxten, R., & Eens, M. (2010). Multiple signals for multiple messages: Great tit, *Parus major*, song signals age and survival. *Animal Behaviour*, 80(3), 451–459. <https://doi.org/10.1016/j.anbehav.2010.06.002>
- Rivera-Gutierrez, H. F., Pinxten, R., & Eens, M. (2011). Difficulties when assessing birdsong learning programmes under field conditions: A re-evaluation of song repertoire flexibility in the great tit. *PLoS One*, 6(1), Article e16003. <https://doi.org/10.1371/journal.pone.0016003>
- Rivera-Gutierrez, H. F., Pinxten, R., & Eens, M. (2012). Tuning and fading voices in songbirds: Age-dependent changes in two acoustic traits across the life span. *Animal Behaviour*, 83(5), 1279–1283. <https://doi.org/10.1016/j.anbehav.2012.03.001>
- Roulin, A. (2001). Food supply differentially affects sibling negotiation and competition in the barn owl (*Tyto alba*). *Behavioral Ecology and Sociobiology*, 49(6), 514–519. <https://doi.org/10.1007/s002650100322>
- Sainburg, T., & Hedley, R. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires.
- Sainburg, T., Theilmann, B., Thielk, M., & Gentner, T. Q. (2019). Parallels in the sequential organization of birdsong and human speech. *Nature Communications*, 10(1). <https://doi.org/10.1038/s41467-019-11605-y>
- Searcy, W. A., Soha, J., Peters, S., & Nowicki, S. (2022). Long-distance dependencies in birdsong syntax. *Proceedings of the Royal Society B: Biological Sciences*, 289(1967), Article 20212473. <https://doi.org/10.1098/rspb.2021.2473>
- Sethi, S. S., Jones, N. S., Fulcher, B. D., Picinali, L., Clink, D. J., Klinck, H., & Ewers, R. M. (2020). Characterizing soundscape across diverse ecosystems using a universal acoustic feature set. *Proceedings of the National Academy of Sciences of the United States of America*, 117(29), 17049–17055. <https://doi.org/10.1073/pnas.2004702117>
- Shabanov, A. (2023). OML-team/open-metric-learning. OML-Team.
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60. <https://doi.org/10.1186/s40537-019-0197-0>
- Sierro, J., de Kort, S. R., & Hartley, I. R. (2023). Sexual selection for both diversity and repetition in birdsong. *Nature Communications*, 14(1), 3600. <https://doi.org/10.1038/s41467-023-39308-5>
- Sierro, J., de Kort, S. R., Riebel, K., & Hartley, I. R. (2022). Female blue tits sing frequently: A sex comparison of occurrence, context, and structure of song. *Behavioral Ecology*, 33(5), 912–925. <https://doi.org/10.1093/beheco/ara044>
- Slagsvold, T., Sætre, G.-P., & Dale, S. (1994). Dawn singing in the great tit (*Parus major*): Mate attraction, mate guarding, or territorial defence? *Behaviour*, 131(1–2), 115–138. <https://doi.org/10.1163/156853994X000244>
- Stowell, D. (2021). Computational bioacoustics with deep learning: A review and roadmap. arXiv:2112.06725 [Cs, Eess, Q-Bio] <http://arxiv.org/abs/2112.06725>.
- Sugai, L. S. M., Silva, T. S. F., Ribeiro, J. W., Jr., & Llusia, D. (2019). Terrestrial passive acoustic monitoring: Review and perspectives. *BioScience*, 69(1), 15–25. <https://doi.org/10.1093/biosci/biy147>
- Thakur, A., Thapar, D., Rajan, P., & Nigam, A. (2019). Deep metric learning for bioacoustic classification: Overcoming training data scarcity using dynamic triplet loss. *Journal of the Acoustical Society of America*, 146(1), 534–547. <https://doi.org/10.1121/1.5118245>
- Thomas, M., Jensen, F. H., Averly, B., Demartsev, V., Manser, M. B., Sainburg, T., & Strandburg-Peshkin, A. (2021). A practical guide for generating unsupervised, spectrogram-based latent space representations of animal vocalizations. Cold Spring Harbor Laboratory. <https://doi.org/10.1101/2021.12.16.472881>
- Thompson, N. S., LeDoux, K., & Moody, K. (1994). A system for describing bird song units. *Bioacoustics*, 5(4), 267–279. <https://doi.org/10.1080/09524622.1994.9753257>
- Thorpe, W. H. (1958). The learning of song patterns by birds, with especial reference to the song of the Chaffinch *Fringilla coelebs*. *Ibis*, 100(4), 535–570. <https://doi.org/10.1111/j.1474-919X.1958.tb07960.x>
- Van Duyse, E., Pinxten, R., Snoeijis, T., & Eens, M. (2005). Simultaneous treatment with an aromatase inhibitor and an anti-androgen decreases the likelihood of dawn song in free-living male great tits, *Parus major*. *Hormones and Behavior*, 48(2), 243–251. <https://doi.org/10.1016/j.yhbeh.2005.02.013>
- van Rossum, G. (1995). *Python reference manual*. CWI.
- Wheelwright, N. T., Swett, M. B., Levin, I. I., Kroodsma, D. E., Freeman-Gallant, C. R., & Williams, H. (2008). The influence of different tutor types on song learning in a natural bird population. *Animal Behaviour*, 75(4), 1479–1493. <https://doi.org/10.1016/j.anbehav.2007.08.030>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1), Article 160018. <https://doi.org/10.1038/sdata.2016.18>

- Williams, H., & Lachlan, R. F. (2021). Evidence for cumulative cultural evolution in bird song. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1843), Article 20200322. <https://doi.org/10.1098/rstb.2020.0322>
- Youngblood, M., & Lahti, D. C. (2022). Content bias in the cultural evolution of house finch song. *Animal Behaviour*, 185, 37–48. <https://doi.org/10.1016/j.anbehav.2021.12.012>
- Zandberg, L., Morfi, V., George, J., Clayton, D. F., Stowell, D., & Lachlan, R. F. (2022). Bird song comparison using deep learning trained from avian perceptual judgments. *bioRxiv*. <https://doi.org/10.1101/2022.12.23.521425>
- Zipple, M. N., Nowicki, S., Searcy, W. A., & Peters, S. (2019). Full life course analysis of birdsong reveals maturation and senescence of highly repeatable song characteristics. *Behavioral Ecology*, 30(6), 1761–1768. <https://doi.org/10.1093/beheco/arz146>

Appendix

There is no consistent terminology used to describe the various hierarchical levels of a bird's vocal production. For clarity, we adopt

the terminology outlined in Thompson et al. (1994); see also Fig. 2b for a graphical explanation.

The fundamental temporal unit is referred to as a note. Notes are represented by continuous traces on the sound spectrogram and are separated by silences. Moving up the hierarchy, syllables are sequences of one or more notes that are always repeated in the same order. Beyond syllables, we have songs, which consist of clusters of the same type of syllables punctuated by longer pauses, often in the order of seconds. Lastly, song bouts are uninterrupted performances of songs of the same type. Great tits tend to sing the same song type repeatedly before transitioning to a different type. They continue this pattern until they stop singing altogether, often after having performed their entire song repertoire.

Table A1

Short description of the files included in the data set

Description
109 963 raw song files with their corresponding metadata
Model-based feature vectors, with 109 963 samples and 384 dimensions
A main derived data set including information on broods, adult bird life history traits, nesting locations and acoustic recordings
Morphological measurements for birds captured and retrapped in the study area. Includes information on species, age, sex, weight and various other traits
Information on the location and characteristics of nestboxes in Wytham Woods, along with a map of the site

For detailed documentation see 'Docs' at <https://nilomr.github.io/great-tit-hits/>.